



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 102649

To: Jennifer Graser

Location: CM1-7E12

Art Unit: 1645

Friday, September 05, 2003

Case Serial Number: 10/049473

From: Beverly Shears

Location: Biotech-Chem Library

CM1-1E05

Phone: 308-4994

beverly.shears@uspto.gov

Search Notes

8/99
Rehoot
Hermanns

98/8930
wo/8/8930
P.55
P114-116
seq. 5021
Interference

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Requestor's

Name: _____

Serial

Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 09-05-03

Searcher: Beverly 29994

Terminal time: 25

Elapsed time: _____

CPU time: _____

Total time: 25

Number of Searches: _____

Number of Databases: 1

Search Site

_____ STIC

_____ CM-1

_____ Pre-S

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

☒ Other CGN

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:33:59 ; Search time 378 Seconds
(without alignments)
741.416 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SNTYKGVDMKKILAGAIR.....IFQYIGGDSSTSSSTSNR 322

Scoring table:
BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 5580241 seqs, 870357830 residues

Total number of hits satisfying chosen parameters: 5580241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending-Patents_AA.Main:*
1: /cgn2_6/ptodata/1/paa/US081.COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US082.COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US083.COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US084.COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US085.COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US086.COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US087.COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US088.COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US089.COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US090.COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US091.COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US092.COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US093.COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US094.COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US095.COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US096.COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US097.COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US098.COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US099.COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US100.COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US101.COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US102.COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US103.COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US104.COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US105.COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US106.COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US107.COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US108.COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US109.COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US110.COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US111.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1588	100.0	322	US-10-049-473a-2	Sequence 2, Appl 1
2	1541	97.0	313	US-09-769-787-163	Sequence 163, App

Interference ARG

3	1538	96.9	313	19	US-09-583-110-5200	Sequence 5200, Ap
4	1428	89.9	295	13 <th>US-09-107-433-4127</th> <th>Sequence 4127, Ap</th>	US-09-107-433-4127	Sequence 4127, Ap
5	916	57.7	213	15 <th>US-08-961-083-34</th> <th>Sequence 34, Appl</th>	US-08-961-083-34	Sequence 34, Appl
6	916	57.7	213	22 <th>US-09-765-271-34</th> <th>Sequence 34, Appl</th>	US-09-765-271-34	Sequence 34, Appl
7	916	57.7	213	22 <th>US-09-765-272-34</th> <th>Sequence 34, Appl</th>	US-09-765-272-34	Sequence 34, Appl
8	916	57.7	213	22 <th>US-09-765-272-34</th> <th>Sequence 34, Appl</th>	US-09-765-272-34	Sequence 34, Appl
9	634	39.9	344	30 <th>US-10-417-884-6355</th> <th>Sequence 6355, Ap</th>	US-10-417-884-6355	Sequence 6355, Ap
10	608.5	38.3	368	15 <th>US-09-134-000-4284</th> <th>Sequence 4284, Ap</th>	US-09-134-000-4284	Sequence 4284, Ap
11	608.5	38.3	368	15 <th>US-09-134-000C-4284</th> <th>Sequence 4284, Ap</th>	US-09-134-000C-4284	Sequence 4284, Ap
12	608.5	38.3	368	15 <th>US-10-434-665-4284</th> <th>Sequence 4284, Ap</th>	US-10-434-665-4284	Sequence 4284, Ap
13	605.5	38.1	342	28 <th>US-10-206-576-182</th> <th>Sequence 182, App</th>	US-10-206-576-182	Sequence 182, App
14	588.5	37.1	303	30 <th>US-10-417-884-6113</th> <th>Sequence 6113, Ap</th>	US-10-417-884-6113	Sequence 6113, Ap
15	562.5	35.4	322	28 <th>US-10-206-576-184</th> <th>Sequence 184, App</th>	US-10-206-576-184	Sequence 184, App
16	450.5	28.4	302	30 <th>US-10-417-884-4736</th> <th>Sequence 4736, Ap</th>	US-10-417-884-4736	Sequence 4736, Ap
17	375.5	23.6	299	26 <th>US-10-049-473a-3</th> <th>Sequence 4736, Ap</th>	US-10-049-473a-3	Sequence 4736, Ap
18	373	23.5	299	26 <th>US-10-049-473a-4</th> <th>Sequence 4736, Ap</th>	US-10-049-473a-4	Sequence 4736, Ap
19	367	22.1	299	26 <th>US-10-049-473a-5</th> <th>Sequence 4736, Ap</th>	US-10-049-473a-5	Sequence 4736, Ap
20	357.5	22.5	292	22 <th>US-09-791-537-115536</th> <th>Sequence 115536, Ap</th>	US-09-791-537-115536	Sequence 115536, Ap
21	342.5	21.6	292	13 <th>US-08-902-393-11</th> <th>Sequence 11, Appl</th>	US-08-902-393-11	Sequence 11, Appl
22	322	20.3	74	1 <th>PCT-US97-07950-508</th> <th>Sequence 508, App</th>	PCT-US97-07950-508	Sequence 508, App
23	294.5	18.5	371	1 <th>PCT-US02-36123-5492</th> <th>Sequence 5492, Ap</th>	PCT-US02-36123-5492	Sequence 5492, Ap
24	234.5	14.8	309	1 <th>PCT-US02-36123-5490</th> <th>Sequence 5490, Ap</th>	PCT-US02-36123-5490	Sequence 5490, Ap
25	217.5	13.7	328	12 <th>US-08-827-355-4943</th> <th>Sequence 4943, Ap</th>	US-08-827-355-4943	Sequence 4943, Ap
26	217.5	13.7	328	20 <th>US-09-611-529-5956</th> <th>Sequence 5956, Ap</th>	US-09-611-529-5956	Sequence 5956, Ap
27	217.5	13.7	328	25 <th>US-09-950-084-5956</th> <th>Sequence 5956, Ap</th>	US-09-950-084-5956	Sequence 5956, Ap
28	212.5	13.4	320	11 <th>US-08-781-986A-5192</th> <th>Sequence 5192, Ap</th>	US-08-781-986A-5192	Sequence 5192, Ap
29	212.5	13.4	320	13 <th>US-08-956-171-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171-5192	Sequence 5192, Ap
30	212.5	13.4	320	13 <th>US-08-956-171B-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171B-5192	Sequence 5192, Ap
31	212.5	13.4	320	13 <th>US-08-956-171C-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171C-5192	Sequence 5192, Ap
32	212.5	13.4	320	13 <th>US-08-956-171D-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171D-5192	Sequence 5192, Ap
33	212.5	13.4	320	13 <th>US-08-956-171E-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171E-5192	Sequence 5192, Ap
34	210	13.2	330	18 <th>US-09-450-969-5963</th> <th>Sequence 5963, Ap</th>	US-09-450-969-5963	Sequence 5963, Ap
35	210	13.2	330	26 <th>US-10-092-411A-3811</th> <th>Sequence 3811, Ap</th>	US-10-092-411A-3811	Sequence 3811, Ap
36	173	10.9	42	22 <th>US-08-832-030-4451</th> <th>Sequence 4451, Ap</th>	US-08-832-030-4451	Sequence 4451, Ap
37	169.5	10.7	594	22 <th>US-09-791-537-146251</th> <th>Sequence 146251, Ap</th>	US-09-791-537-146251	Sequence 146251, Ap
38	163	10.3	621	23 <th>US-09-897-516-7072</th> <th>Sequence 7072, Ap</th>	US-09-897-516-7072	Sequence 7072, Ap
39	163	10.3	621	31 <th>US-06-215-161-7072</th> <th>Sequence 7072, Ap</th>	US-06-215-161-7072	Sequence 7072, Ap
40	159	10.0	677	16 <th>US-09-252-691-8092</th> <th>Sequence 8092, Ap</th>	US-09-252-691-8092	Sequence 8092, Ap
41	159	10.0	677	16 <th>US-09-252-691C-8092</th> <th>Sequence 8092, Ap</th>	US-09-252-691C-8092	Sequence 8092, Ap
42	159	10.0	677	30 <th>US-10-417-886-8092</th> <th>Sequence 8092, Ap</th>	US-10-417-886-8092	Sequence 8092, Ap
43	155.5	9.8	282	22 <th>US-09-791-537-120537</th> <th>Sequence 120537, Ap</th>	US-09-791-537-120537	Sequence 120537, Ap
44	155.5	9.8	336	22 <th>US-09-791-537-39861</th> <th>Sequence 39861, A</th>	US-09-791-537-39861	Sequence 39861, A
45	147.5	9.3	447	30 <th>US-10-419-128-17422</th> <th>Sequence 17422, A</th>	US-10-419-128-17422	Sequence 17422, A

ALIGNMENTS

RESULT 1
US-10-049-473a-2
Sequence 2, Application US/10049473A
GENERAL INFORMATION:
APPLICANT: de Groot, Ronald
TITLE OF INVENTION: Pneumococcal Vaccines
FILE REFERENCE: Docket 294-120 PCT/US
CURRENT APPLICATION NUMBER: US/10/049, 473A
CURRENT FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: PCT/NL00/00569
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: EP 99202640.1
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 322
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-049-473a-2
Query Match 100.0%; Score 1588; DB 26; Length 322;
Best Local Similarity 100.0%; Pred. No. 1,4e-113;
Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SNTYKGVDMKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP 60
| | | | |
Db 1 SNTYKGVDMKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP 60
QY 61 SAQOVLNMTIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQ 120
| | | | |
Db 61 SAQOVLNMTIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQ 120
QY 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
| | | | |
Db 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
QY 181 DEAOAKNSDDEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYS 240
| | | | |
Db 181 DEAOAKNSDDEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYS 240
QY 241 SOYIVLKLTKTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAF 300
| | | | |
Db 241 SOYIVLKLTKTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAF 300
QY 301 QNIFQYIGGGDSSSSSTSNE 322
| | | | |
Db 301 QNIFQYIGGGDSSSSSTSNE 322

RESULT 2
US-09-769-787-163

Sequence 163, Application US/09769787
GENERAL INFORMATION:
APPLICANT: Microbial Techniques Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Phillip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PMC/P21129WO
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 163
LENGTH: 313
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-163

Query Match 97.0%; Score 1541; DB 22; Length 313;
Best Local Similarity 100.0%; Pred. No. 5, 6e-110;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 MKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP5AQOVLN 69
| | | | |
Db 1 MKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP5AQOVLN 69
QY 70 TIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQITSLVEL 129
| | | | |
Db 61 TIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQITSLVEL 129
QY 130 AVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGADPAQAKDN 189
| | | | |
Db 121 AVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGADPAQAKDN 189
QY 190 STEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYSQYIVLKL 249
| | | | |
Db 181 STEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYSQYIVLKL 249
QY 250 KTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAFONIFQYIG 309
| | | | |
Db 241 KTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAFONIFQYIG 309

QY 310 GGDSSSSSTSNE 322
| | | | |
Db 301 GGDSSSSSTSNE 313

RESULT 3
US-09-583-110-5200

Sequence 5200, Application US/09583110
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PAT00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 5200
LENGTH: 313
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-5200

Query Match 96.9%; Score 1538; DB 19; Length 313;
Best Local Similarity 99.7%; Pred. No. 9, 5e-110;
Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 10 MKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP5AQOVLN 69
| | | | |
Db 1 MKKLLAGATLLSVATLAACSGSGADLISMKGVITHEHOFYEQVKSNP5AQOVLN 69
QY 70 TIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQITSLVEL 129
| | | | |
Db 61 TIOKVEKOGSELDKEDVDITAEKKOYGENYOVL5OAGMTLETTRAAQITSLVEL 129
QY 130 AVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGADPAQAKDN 189
| | | | |
Db 121 AVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGADPAQAKDN 189
QY 190 STEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYSQYIVLKL 249
| | | | |
Db 181 STEKRENGEITFDASSTEVEPOVKAALDVGSDVITATGTQAYSQYIVLKL 249
QY 250 KTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAFONIFQYIG 309
| | | | |
Db 241 KTEKSSNIDYKREKLTIVILQKONDSTFVOSIIGKELOANIKVKDQAFONIFQYIG 309
QY 310 GGDSSSSSTSNE 322
| | | | |
Db 301 GGDSSSSSTSNE 313

RESULT 4
US-09-107-433-4127

Sequence 4127, Application US/09107433
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

```

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arianello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4127:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc-feature
LOCATION: (B) LOCATION 1...295
SEQUENCE DESCRIPTION: SEQ ID NO: 4127:
US-09-107-433-4127

```

Query Match 89.9%; Score 1428; DB 15; Length 295;
 Best Local Similarity 99.3%; Pred. No. 2,6e-101;
 Matches 289; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 32 KSEBADIISMKGDIYTHQFYEQVKSNSAQOVLNMTIOKVFKEQYSGSELDKREYDT 91
DB 5 KSEBADIISMKGDIYTHQFYEQVKSNSAQOVLNMTIOKVFKEQYSGSELDKREYDT 64
QY 92 IAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVELAVKVAEAEITDEAYKKAFDEY 151
DB 65 IAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVELAVKVAEAEITDEAYKKAFDEY 124
QY 152 TPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTEV 211
DB 125 TPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTEV 184
QY 212 PPOVKAFAALVDVSDYITATGTAQYSSQYIYKLTTRKTEKSSNIDYKREKLVIT 271
DB 185 PPOVKAFAALVDVSDYITATGTAQYSSQYIYKLTTRKTEKSSNIDYKREKLVIT 244
QY 272 QKQNSTFVQSIIGKEIKELQANIKVQDAFQNIPTQYIGGSDSSSSSTSSNE 322
DB 245 QKQNSTFVQSIIGKEIKELQANIKVQDAFQNIPTQYIGGSDSSSSSTSSNE 295

```

RESULT 5
 US-08-961-083-34
 Sequence 34, Application US/08961083
 GENERAL INFORMATION:
 APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA

```

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-08-961-083-34

```

Query Match 57.7%; Score 916; DB 13; Length 213;
 Best Local Similarity 93.6%; Pred. No. 4.5e-62;
 Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

```

QY 31 SKGSEBADIISMKGDIYTHQFYEQVKSNSAQOVLNMTIOKVFKEQYSGSELDKREYD 90
DB 1 SKGSEBADIISMKGDIYTHQFYEQVKSNSAQOVLNMTIOKVFKEQYSGSELDKREYD 60
QY 91 TIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVELAVKVAEAEITDEAYKKAFDE 150
DB 61 TIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVELAVKVAEAEITDEAYKKAFDE 120
QY 151 TTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTEV 210
DB 121 TTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTEV 180
QY 211 VP-EQVKAFA-----LDVD 225
DB 181 VPGASPKPLFAFCGMVFLDVD 203

```

RESULT 6
 US-09-765-271-34
 Sequence 34, Application US/09765271
 GENERAL INFORMATION:
 APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/765,271
 FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/536,784
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-765-271-34

Query Match 57.7%; Score 916; DB 22; Length 213;
Best Local Similarity 93.6%; Pred. No. 4.5e-62;

Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

QY 31 SKSGEGADLISKMGDVTTEHOFEYQVKSNSPSAQVLLNMTIOKVEKQYSELDDEVD 90
DB 1 SKSGEGADLISKMGDVTTEHOFEYQVKSNSPSAQVLLNMTIOKVEKQYSELDDEVD 60
QY 91 TIAEEKQYGENYQVLSQAGMTLETTRKAQIRTSKVLAVKVAEALTDYAKKAFDE 150
DB 61 TIAEEKQYGENYQVLSQAGMTLETTRKAQIRTSKVLAVKVAEALTDYAKKAFDE 120
QY 151 YTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTDEKTKENGSEITPDSASTE 210
DB 121 YTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTDEKTKENGSEITPDSASTE 180
QY 211 VP-EQVKAFA-----LDVD 225
DB 181 VPGASPKPLFAFRGMYFLVD 203

RESULT 7

US-09-765-272-34

Sequence 34, Application US/09765272

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-765-272-34

Query Match 57.7%; Score 916; DB 22; Length 213;
Best Local Similarity 93.6%; Pred. No. 4.5e-62;

Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

QY 31 SKSGEGADLISKMGDVTTEHOFEYQVKSNSPSAQVLLNMTIOKVEKQYSELDDEVD 90
DB 1 SKSGEGADLISKMGDVTTEHOFEYQVKSNSPSAQVLLNMTIOKVEKQYSELDDEVD 60
QY 91 TIAEEKQYGENYQVLSQAGMTLETTRKAQIRTSKVLAVKVAEALTDYAKKAFDE 150
DB 61 TIAEEKQYGENYQVLSQAGMTLETTRKAQIRTSKVLAVKVAEALTDYAKKAFDE 120
QY 151 YTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTDEKTKENGSEITPDSASTE 210
DB 121 YTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTDEKTKENGSEITPDSASTE 180
QY 211 VP-EQVKAFA-----LDVD 225
DB 181 VPGASPKPLFAFRGMYFLVD 203

RESULT 8

US-09-765-272A-34

Sequence 34, Application US/09765272A

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and

NUMBER OF SEQUENCES: 454

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: Dell Latitude C610

OPERATING SYSTEM: Windows 2000

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Lin J. Hymel

REGISTRATION NUMBER: 45,414

REFERENCE/DOCKET NUMBER: PB340P2C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 610-5790

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 213 amino acids

TYPE: amino acid


```

Db      262 KMTNKAAGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
      307 YIGGDSSTSSSTSSNE 322
      322 YM-QTESSASSEKKE 336

```

```

RESULT 11
US-09-134-000C-4284
; Sequence 4284, Application US/09134000C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4284
; LENGTH: 368
; TYPE: PRM
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4284

```

```

Query Match      38.3%; Score 608.5; DB 15; Length 368;
Best Local Similarity 41.8%; Pred. No. 5e-38;
Matches 132; Conservative 63; Mismatches 112; Indels 9; Gaps 5;

```

```

OY      10 MKKILAGATILLSATLACSGSEGADLISMKGVITEHQFVEQVSNPSAQVLLNM 69
      27 MKKILILAAAGAMVFLAACSSGSK--DIATKMGSTIYDDFYNQIKQSTSQQAFSQM 84
      70 TIQKVEKQVSELDKEDVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 126
      85 VYKVEEYKDYDKDIOKNFDEAKQVPAQGGKSDLKQAGLTKETFKQKORAPA 144
      127 VELAVKVAEALDEAVKKAFADEYTPDVTQAQIRLNNEDEKAEVLEKAKAEGADFAOLA 186
      145 YDAGLK--AHLKTTDEDLKTAMASFHPEVEAQIIQVASEDDAKAV--KKEITDGDFTKIA 201
      187 KDNSTDEKTEKNGEITFDSASTEVPEQVKAFAALDVGVSDVITATGTQAVSSQYIY 246
      202 KEKSTDTATKKGDKIKFDSQATVPAEYKKAFAKLDGEVSEPIAATNMQTYOTTYV 261
      247 KTKKTEKSSNIDYKKEIKTYILTKQKNDSTFVOSIIGKELQANIKVKDQAFQNIPTQ 306
      262 KMTNKAAGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
      307 YIGGDSSTSSSTSSNE 322
      322 YM-QTESSASSEKKE 336

```

```

RESULT 12
US-10-434-665-4284
; Sequence 4284, Application US/10434665
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-09
; CURRENT APPLICATION NUMBER: US/10/434,665
; CURRENT FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US 09/134,000
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 4284
; LENGTH: 368
; TYPE: PRM
; ORGANISM: Enterococcus faecalis
US-10-434-665-4284

```

```

Query Match      38.3%; Score 608.5; DB 30; Length 368;
Best Local Similarity 41.8%; Pred. No. 5e-38;
Matches 132; Conservative 63; Mismatches 112; Indels 9; Gaps 5;

```

```

OY      10 MKKILAGATILLSATLACSGSEGADLISMKGVITEHQFVEQVSNPSAQVLLNM 69
      27 MKKILILAAAGAMVFLAACSSGSK--DIATKMGSTIYDDFYNQIKQSTSQQAFSQM 84
      70 TIQKVEKQVSELDKEDVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 126
      85 VYKVEEYKDYDKDIOKNFDEAKQVPAQGGKSDLKQAGLTKETFKQKORAPA 144
      127 VELAVKVAEALDEAVKKAFADEYTPDVTQAQIRLNNEDEKAEVLEKAKAEGADFAOLA 186
      145 YDAGLK--AHLKTTDEDLKTAMASFHPEVEAQIIQVASEDDAKAV--KKEITDGDFTKIA 201
      187 KDNSTDEKTEKNGEITFDSASTEVPEQVKAFAALDVGVSDVITATGTQAVSSQYIY 246
      202 KEKSTDTATKKGDKIKFDSQATVPAEYKKAFAKLDGEVSEPIAATNMQTYOTTYV 261
      247 KTKKTEKSSNIDYKKEIKTYILTKQKNDSTFVOSIIGKELQANIKVKDQAFQNIPTQ 306
      262 KMTNKAAGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
      307 YIGGDSSTSSSTSSNE 322
      322 YM-QTESSASSEKKE 336

```

```

RESULT 13
US-10-206-576-182
; Sequence 182, Application US/10206576
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: P3369P1D1
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:

```

```

?      LENGTH: 342 amino acids
?      type: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
?      SEQUENCE DESCRIPTION: SEQ ID NO: 182
?
HS-10-206-576-182

```

Query Match	38.1%	Score 605.5;	DB 28;	Length 342;
Best Local Similarity	41.8%	Pred. No. 7.7e-38;		
Matches 132;	Conservative 62;	Mismatches 113;	Indels 9;	Gaps 5

Qy	10	MKKKLLAGAITLLSVATLLACSKSGEADLLMSKEDVYTHHQFPEQVSNQAOVLNM	69
		: : : : : : : : : : : : : : : : : : : :	
Db	1	MKKKLLAAGAAMAVPSLACSSGSK--DIATMKSTITVYDFNQIKEDSTSQANSSQM	58
Qy	70	TIOKVERKQSGSLDDKEVDOTIAEKKO---YGNTRYVLSQAGMLTFPKAOIRISKL	126
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	59	VITKVEEKYGGDYVDDKXIKQKNPDEAKGVEAOGSKFSDAKOAGLETKEFKKLOKRAA	118
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	127	VELAVKVAEAEITDEAYKKAFDEYTPDTAQIITLNNEDRAKEVLEKAKEGADFAOLA	186
Qy		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	119	YDAGLK--AHLKITDEDTLKTAMASHPEVEAQIIVASEDDAKAV-KKEITDGGDFPKIA	175
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	187	KDNSTDEKTRKENGGETTFPDSASTEYPEQYKKAAFLVDVGSVYITATGTQAYSQYIIV	246
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	176	KEKSTDTAKKDDGKIKTFDSQATTPYPAEKVEKAFLKDGVESEPTAAANMOTYPTTYVV	235
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	247	KLAKKTEKSSNIDYDEKELKTAYILOKONDSTFYVSIIGKELOANIKVKKDOAQNIFTQ	306
Qy		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	236	KMTKRNKAKGNDMPYKEIKKITAEPFKLADQFFVSKVISDELKAANVKKINDADKNNLAG	295
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	307	YIGGGSSSSSSSTSTNE 322	
Db		: : : : : : : : : : : : : : : : : : : : : : : : :	
		296 YM-QTESSSSASSEKKE 310	

RESULT 14
 US-10-417-884-6113
 Sequence 6113, Application US/10417884
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 City: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/417.884
 FILING DATE: 17-Apr-2003
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/107.532A
 FILING DATE: 30-Jun-1998
 APPLICATION NUMBER: 60/085.598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinielello, Pamela Deneké
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 6113:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 303 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEetical: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (B) LOCATION 1...303
 SEQUENCE DESCRIPTION: SEQ ID NO: 6113
 US-10-417-884-6113

Query Match	37.1%	Score 588.5	DB 30	Length 303
Best Local Similarity	40.5%	Pred. No. 1.3e-36		
Matches 125	Conservative 69	Mismatches 108	Indels 7	Gaps 5

[illegible]

```

1 RESULT 15
2 US-10-206-576-184
3 ; Sequence 184, Application US/10206576
4 ; GENERAL INFORMATION:
5 ; APPLICANT: Choi et al.
6 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
7 ; NUMBER OF SEQUENCES: 497
8 ; CORRESPONDENCE ADDRESSES:
9 ; ADDRESSEE: Human Genome Sciences, Inc.
10 ; STREET: 9410 Key West Avenue
11 ; CITY: Rockville
12 ; STATE: Maryland
13 ; COUNTRY: USA
14 ; ZIP: 20850
15 ;
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: CD-R
18 ; COMPUTER: Dell Latitude
19 ; OPERATING SYSTEM: Windows 98
20 ; SOFTWARE: ASCII Text
21 ;
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/10/206,576
24 ; FILING DATE: 29-Jul-2002
25 ; CLASSIFICATION: <Unknown>
26 ;
27 ; PRIOR APPLICATION DATA:
28 ; APPLICATION NUMBER: US 09/071,035
29 ; FILING DATE: 1998-05-04
30 ; APPLICATION NUMBER: US 60/046,655
31 ; FILING DATE: 1997-05-16
32 ;

```

Fri Sep 5 09:24:50 2003

APPLICATION NUMBER: US 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: US 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Hyman, Mark J.
 REGISTRATION/DOCKET NUMBER: PB369P1D1
 INFORMATION FOR SEQ ID NO: 184:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 322 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 184:
 US-10-206-576-184

Query Match 35.48; Score 562.5; DB 28; Length 322;
 Best Local Similarity 41.2%; Pred. No. 1,5e-34;
 Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;
 30 CSNGSGADLISMKGSDVITEHQYEQVKSNPFAQVILNMTIQVFEKQYGSLEDDKEVD 89
 1 CSSGSK--DIATMKGSTIVDDFYNOIKEOSTSOAFSOMVIYKVFEEKYGDKVDKXIO 58
 90 DTIAEKKQ---YGENYORVLSOAGMTLETRKAQIRTSKVELAVKVAEALTDEAYKK 146
 59 KNFDEAKEOVEAOGKFSQALMKGALTEKTFKQIKORAAVDAGLK--AHLKITDEDLKT 116
 147 AFDEYTPDVTAQIIRLNNEBKAEVLEKAKAGADFAQLAKDNSTDEKTKENGSETFDS 206
 117 AMASFPHEVEKQIIOVASSEDAAKAV--KKEITDGGDTLAKKSTDTATKKGKIKEDS 175
 207 ASTEYPEQYKKAFAFDVDSVITATGTQAVSSQYIVKLTKEKSSNIDYKEXLK 266
 176 QATVPAYKKAFAFKLKGVESEPIAATNMQTOTIYVYKTKKAKGNDMKPYEKIK 235
 267 TVILTKQKNDSTFVQSTIGKELQANIKVRDQAFONITQYIGGGDSSSSSTSN 322
 236 KIAEETKILADQTFVSKVISDELKAAVYKIKDDAFKNALAGYM-QTESSSSASSEKKE 290

Search completed: August 29, 2003, 10:44:00
 Job time : 380 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:34:29 ; Search time 25 Seconds
(without alignments)
381.116 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SNTYLKGVDMKKILAGAIR.....IFTQYIGGDSSTSSSTSSNE 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141812 seqs, 29589763 residues

Total number of hits satisfying chosen parameters: 141812

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1428	89.9	295	US-10-617-320-4127	Sequence 4127, Ap
2	163	10.3	621	US-09-897-516A-7078	Sequence 7078, Ap
3	129.5	8.2	589	US-10-299-636-97	Sequence 97, Appl
4	129.5	8.2	929	US-10-299-636-94	Sequence 94, Appl
5	129.5	8.2	1314	US-60-487-610-2694	Sequence 2694, Ap
6	127.5	8.0	1790	US-10-326-956-1291	Sequence 1291, Ap
7	126	7.9	643	US-10-299-636-95	Sequence 95, Appl
8	123.5	7.8	630	US-10-603-113-20275	Sequence 20275, A
9	123	7.7	588	US-10-299-636-96	Sequence 96, Appl
10	122	7.7	901	US-10-603-113-14747	Sequence 14747, A
11	121.5	7.7	802	US-10-273-573-9033	Sequence 9033, Ap
12	121.5	7.7	1154	US-10-273-573-9040	Sequence 9040, Ap
13	119	7.5	336	US-10-299-636-103	Sequence 103, App
14	119	7.5	354	US-10-299-636-105	Sequence 105, App
15	119	7.5	713	US-10-257-377-3	Sequence 3, Appl
16	118.5	7.5	2101	US-09-286-6628-32	Sequence 32, Appl
17	118.5	7.5	2207	US-10-273-573-9039	Sequence 9039, Ap
18	117.5	7.4	200	US-10-299-636-50	Sequence 50, Appl
19	117.5	7.4	2115	US-09-296-6628-33	Sequence 33, Appl
20	117.5	7.4	2121	US-10-273-573-9034	Sequence 9034, Ap
21	116.5	7.3	301	US-10-603-113-26282	Sequence 26282, A
22	116.5	7.3	1875	US-10-326-956-2369	Sequence 2369, Ap
23	116	7.3	979	US-10-603-114-5466	Sequence 5466, Ap
24	116	7.3	1014	PCT-US03-21379-81	Sequence 81, Appl
25	116	7.3	1960	US-10-293-244-1516	Sequence 1516, Ap
26	116	7.3	1963	US-10-293-244-3484	Sequence 3484, Ap

27	115.5	7.3	299	US-09-065-756B-61	Sequence 61, Appl
28	115.5	7.3	448	US-10-603-113-20279	Sequence 20279, A
29	115.5	7.3	635	US-10-603-114-5956	Sequence 5956, Ap
30	115	7.2	1988	US-10-286-897-5930	Sequence 5930, Ap
31	115	7.2	1988	US-10-286-897-5931	Sequence 5931, Ap
32	115	7.2	1988	US-10-258-898A-5930	Sequence 5930, Ap
33	115	7.2	1988	US-10-258-898A-5931	Sequence 5931, Ap
34	114.5	7.2	2117	US-10-273-573-9038	Sequence 9038, Ap
35	114	7.2	639	US-10-603-113-20583	Sequence 20583, A
36	112.5	7.1	206	US-10-603-113-20583	Sequence 20583, A
37	112.5	7.1	1341	US-10-603-113-19074	Sequence 19074, Ap
38	111.5	7.0	487	US-10-333-120A-9	Sequence 9, Appl
39	111	7.0	428	PCT-US02-34769-8	Sequence 8, Appl
40	111	7.0	429	PCT-US02-34769-16	Sequence 16, Appl
41	111	7.0	1213	US-10-286-897-1161	Sequence 1161, Ap
42	111	7.0	1213	US-10-258-898A-3161	Sequence 3161, Ap
43	110.5	7.0	299	US-09-065-756B-341	Sequence 341, App
44	110.5	7.0	417	US-60-487-610-1535	Sequence 1535, Ap
45	110.5	7.0	427	US-10-326-956-2877	Sequence 2877, Ap

ALIGNMENTS

RESULT 1
US-10-617-320-4127
Sequence 4127, Application US/10617320
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIA
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESSES:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariellello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ. ID NO: 4127:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1...295
SEQUENCE DESCRIPTION: SEQ ID NO: 4127
US-10-617-320-4127

Query Match 89.9%; Score 1428; DB 6; Length 295;
Best Local Similarity 99.3%; Pred. No. 6e-93;
Matches 289; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 32 KGSAGALLSMKGVITHEHOFEYEVKSNPSAQOVLNMTIOKVEKQYSGELDDKEDDPT 91
DB 5 KVSAGALLSMKGVITHEHOFEYEVKSNPSAQOVLNMTIOKVEKQYSGELDDKEDDPT 64
QY 92 IAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVKVAEALTEAAYKKADEX 151
DB 65 IAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVKVAEALTEAAYKKADEX 124
QY 152 TPQVTAQIRLNNDKAKVELEKAKAGADFAQAKNSNDEKTEKNGGETTDSASTEV 211
DB 125 TPQVTAQIRLNNDKAKVELEKAKAGADFAQAKNSNDEKTEKNGGETTDSASTEV 184
QY 212 PEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLTKEKSSNIDYKKEKLTIVLT 271
DB 185 PEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLTKEKSSNIDYKKEKLTIVLT 244
QY 272 OKQNDSTFVQSIIGKELQAAANIKVKDQAFONITQYIGGDDSSSSSTSSNE 322
DB 245 OKQNDSTFVQSIIGKELQAAANIKVKDQAFONITQYIGGDDSSSSSTSSNE 295

RESULT 2

US-09-897-516a-7078
Sequence 7078, Application US/0989/516A
GENERAL INFORMATION:
APPLICANT: Corbin, David R.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Huesing, Joseph E.
APPLICANT: Malvar, Thomas M.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Slater, Steven C.
APPLICANT: Spiridonov, Sergei
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
FILE REFERENCE: 38-21(51847)B
CURRENT APPLICATION NUMBER: US/09/897, 516A
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215,161
PRIOR FILING DATE: 2000-06-30
NUMBER OF SEQ ID NOS: 8415
SEQ ID NO 7078
LENGTH: 621
TYPE: PRT
ORGANISM: Xenorhabdus sp.
US-09-897-516a-7078

Query Match 10.3%; Score 163; DB 5; Length 621;
Best Local Similarity 20.4%; Pred. No. 0.00042;
Matches 92; Conservative 72; Mismatches 136; Indels 152; Gaps 19;

QY 13 KLGAATLLSVATLAACSGSEGADLIS--MKGVITE--HQFEYQV----- 57
DB 15 KIVALLILITFLVGLTGYLSSEGNYAAKNGQISRAQLEQVFOQDKMSLOERLDOF 74
QY 58 ----SNPSAQOVL--LNMTIOKVEKQYSGELDDKEDDPTIAEE--KKQY----- 99
DB 75 SALLSNEQEVRAALKRQSDSIINTVLLLEOYANKLGLSASDQVKEIRKQRYFQTDGKFD 134
QY 100 GENYQVLSQAGMTLETTRKAQI-----RTSK 125
DB 135 NEKTIQVITNAGLNDPNAEQRIDRDLISROLITIMSTELPAEIKQALQSLQERKAR 194
QY 126 LVEALAKV--AEALTEDEA----- 143
DB 195 LATLELKSIEAQSVTDQELKNYVNMNSKHFTVEPKVKSIIYIKMDADELKNVTVDADI 254

QY 144 ---YKKADEYTPDVTAO--IIRLNNDKAKVELEKAKAGADFAQAKNSNDEKTEKEN 198
DB 255 EKRYKNNLSKTYTPPEKKYSLSIOLASEADAKSVLDLK--KGADFSLAEKSTDKSAQO 313
QY 199 GGEITPDSASTEVPEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLT--KTEKSSN 257
DB 314 GGDIGWMEESS--PDEI--KSAHLTEKQOLSAVI-----KLSNGYAVFRLDIDIKPQVVP 365
QY 258 IDDKKELKTVILITQKONSTFV-----OSIIGKELQAAANIKV----- 295
DB 366 LADERSALETQVQOEALIDFVALQKRVSEAAANDNESLAAE--QAACTKAYTTDMFDD 424
QY 296 ---KQOAFONITQYIGG--DSSSSSTSSNE 322
DB 425 HVPADINFSQYVQATFFGNLVDDKSPSGTNSD 456

RESULT 3

US-10-299-636-97
Sequence 97, Application US/10299636
GENERAL INFORMATION:
APPLICANT: Briles, David E
APPLICANT: McDaniel, Larry S
APPLICANT: Swaello, Edwin
APPLICANT: Yohef, Janet
APPLICANT: Crain, Marilyn J
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
FILE REFERENCE: 57909/361
CURRENT APPLICATION NUMBER: US/10/299, 636
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: 08/714,741
PRIOR FILING DATE: 1996-09-16
PRIOR APPLICATION NUMBER: 08/529,055
PRIOR FILING DATE: 1995-09-15
NUMBER OF SEQ ID NOS: 111
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 97
LENGTH: 589
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-299-636-97

Query Match 8.2%; Score 129.5; DB 6; Length 589;
Best Local Similarity 22.5%; Pred. No. 0.087;
Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGVITEHOFEYEVKSNPSA--QOVL--LNMTIOKVEKQYSGELDDKEDDPTIAEE 95
DB 232 VSGELATPDKRENDKSSDSSVGEETLPSPSLNANSSQTEHR-----KQVDEYI--- 281
QY 96 KQYGE--NYQVLSQAGMTLETTRKAQIRTSKLVELAV----- 131
DB 282 KKMSELQDRKRHTQOVNINIKISAIKTYIVELSVLKNSKKELTSTKALTEFAFE 341
QY 132 -----KVAEAL--LIDEAYKKADEYTPD----- 154
DB 342 QFKKDTLKPEKVAEAKVEAKKAKKADQKEDRRNYPNTYKTELEIAESDVYKKA 401
QY 155 ----VTQOIRLNNDKAKVELEKAKAGADFAQAKNSNDEKTEKNGGETTDSASTE 210
DB 402 ELELVKEANESRNEEKIKQAKVESEKKAETPLER--ITDRKKA-----EAKRK 453
QY 211 VPEQYKKAFAFD--VDSVSDVITATGTQAVSSQYIVKLTKEK--SSNIDY--KEKLR 266
DB 454 AESEKKAAREKQKQVDAEVALAE--KIALEVEQVRLEKELKEIDESSEDTYKRGRL 510
QY 267 TVI--ITQKONSTFVQSIIGK-----ELQAAANIKVKDQAFONITQYIGGDDSSSS 317
DB 511 APLQSKIDTKKAKLSKLEELSDKIDELDAETIAKLEVLQKDAEGNNNVEAYFEGLEKTTA 570

QY 318 STSNE 322
Db 571 EKRAE 575

RESULT 4

US-10-299-636-94
; Sequence 94, Application US/10299636
; GENERAL INFORMATION:
; APPLICANT: Billes, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swatlo, Edwin
; APPLICANT: Yocher, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-94

Query Match 8.2%; Score 129.5; DB 6; Length 929;
Best Local Similarity 22.5%; Pred. No. 0.15;

Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGDVITEHOFYEQVKNPSA--OOVL---LNMTOQVFEKQYGSLEDDKEVDYTAEE 95
Db 269 VSGELATPPKKNENDAKSSDSVGEETLPSPLNNANESQTEHR-----KDVEYI--- 318
QY 96 KROYGE-NYORVLSQAGMTLETRKAQIRTSKLYELAV----- 131
Db 319 KKMUSEIQDRRHQTQVNNIKLSAIKTYIVELSLVLEKNSKKEELTSKKAELTAFFE 378
QY 132 -----KKVAEAE-LTDEAYKKAPEDEYTPD----- 154
Db 379 QFKKDTLPKPKYVAEAEKKVEAKKKAKKQKEEDRRNYPTNTYKTELEIAESDVKYEA 438
QY 155 -----YTAQIIRLNNEDAKKEVLEKAKAGDFQOLAKDSTDEKTKENGGEITPDSASTE 210
Db 439 ELEVKEANESRREKEIKQAKKEVESKKAATRLER-ITDRKKAE-----EEAKRK 490
QY 211 VPEQVKKAFAALD-VDSGVDTATGTQAVSSQYIYIKTKTEK--SSNIDY-KKKLK 266
Db 491 AESEKKAABAKQKVDAEYALEA---KIAELEVEYGRLEKELKEIDESDSEDLKGLR 547
QY 267 TVI---LTOKONDSTFVQSIIGK-----ELQANIKVKKDAFONITFYIYGDDSSSS 317
Db 548 APLQSKLDTRKAKISLEELSDKIDELDAEIAKLEVLKDAEGNNNEAVEAFKEGLEKTTA 607
QY 318 STSNE 322
Db 608 EKRAE 612

RESULT 5

US-60-487-610-2694
; Sequence 2694, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: C0001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2694
; LENGTH: 1314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-487-610-2694

Query Match 8.2%; Score 129.5; DB 7; Length 1314;
Best Local Similarity 20.9%; Pred. No. 0.24;

Matches 65; Conservative 73; Mismatches 136; Indels 37; Gaps 13;

QY 32 KSGEADLISMKG---DVITEHOFYEQVKNPSAQVILNMTIOKVEKQYGSLEDDK-- 86
Db 753 KQLEESLNLNLGTDGDMYTDQAFEDRLK---EAREYMDLLREAOVKKDQNTLMDRLO 809
QY 87 EVDDTIAEEKQYGENYQVLSQAGMTLETRKAQI-RTSKVLVLAIVKKAABELTDEAYK 145
Db 810 RVNNTLSSQISRL-QNIRNTIEETGNLAEQARAHVENTERLIEIASHELEKAKVAANVS 868
QY 146 KAPEYTPDYTAQIIRLNN-----EDKAKEVLEKAKAGADPAOLAK---DNSTDE----- 193
Db 869 VTQPESTGDP-----NNMTLLAEFAKKLAERHQADIVRAKANDSTGEAYVILL 921
QY 194 KTKENGGEITFD-SASTVEPEYQVKAFAALDVG--VSDVITATGTOA---YSSQYIYK 247
Db 922 RTLAGENQTAPELIELRLKKEEQAKNISQDEKQAARHEAKRAGDAVELIYASVQLOSP 981
QY 248 LTKRT--EKSNNIDYKREKLYILTKQONDSTFVQSIIGKELOANIKVKKDAFONIFT 305
Db 982 LDSETLNENANNIKMEANLEQLIDKLYEDLRDMRGELEVKMLKKGTEQQTAD 1041
QY 306 QYICGGSSSS 316
Db 1042 QILARADAAKA 1052

RESULT 6

US-10-326-956-1291
; Sequence 1291, Application US/10326956
; GENERAL INFORMATION:
; APPLICANT: Bauer et al.
; TITLE OF INVENTION: Protein Complexes and Methods for their use
; FILE REFERENCE: 220615
; CURRENT APPLICATION NUMBER: US/10/326,956
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: EP 01 130 253.6
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 3282
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1291
; LENGTH: 1790
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-326-956-1291

Query Match 8.0%; Score 127.5; DB 6; Length 1790;
Best Local Similarity 22.8%; Pred. No. 0.48;

Matches 69; Conservative 59; Mismatches 103; Indels 71; Gaps 15;

QY 49 EHOFEYEQV---SNPSAQVILNMTIOKVEKQY-----SELDKEVD--D 90
Db 1172 EROYNEISQUNDITSTQGE--NESIKKNNDELGEVKKMKSTSEQSNLUKKSSEIDALN 1229
QY 91 TIAEEKQYGENYQVLSQAGMTLETRKAQI-----RTSKVLVLAIVK----- 132
Db 1230 LQIKELKKKNETNASLSEISIKVSESETVIKELQEDCNFKEKVESLEDEKLKASEDKNS 1289

	Query Match	7.9%;	Score 126;	DB 6;	Length 643:
	Best Local Similarity	22.8%;	Pred.	No. 0.17,	
	Matches	79;	Conservative	Mismatches 117;	Indels 104; Gaps 15.
OY	54	EQVNSPSAQOVLNMNTIQKVFEEKQYGSELDDEKVEDDTTAEKKQGE-NYORVLSQAQM	112		
Dd	7	BETLSPSP-----LNMANESQTEHR-----KDVEYI---KKMSLEIQIDRRKHQTQAV	51		
OY	113	TLETAKAIRTSTKIYELAV-----	138		
Dd	52	NINIKLSAIKKRYVELSVLKENSKEEITSKTAELTAFAEQFKDKTLTPERKVAEAEEK	111		
OY	139	.LTDEAYKKAFFDEYPD-----	168		
Dd	112	KVEEAKKAKKOQKEDDRNYPNTNYKTLELELESQDVYKKAELAYLEYKEEANESRREKET	171		
OY	169	KEVLEKAKABEGADFAQLAKNDSTDEKTKENGGEITPDSASTEVPEQVKRAAFALD-VDGV	227		
Dd	172	KQAKEVESKKAEMTRLEK-IKTDRKAAE-----EAKKRAEEBEKKAAAKAKQVDAE	223		
OY	228	SDVIATAGTQAVSSQYIVKLTKTEK--SSNIDY-KKKLTVI--LTQRONDSTFVQ	281		
Dd	224	EYAIEA---KIAELEVEYORLEKELKEIDESDESDEYLKGLAPLOSKLDTKAKTLSTLE	280		
OY	1282	SIIKG-----ELQANIKYKDOAFQIFITQYIGGDDSSSSSTSNE	322		
Dd	261	ELSKRIDELDAELAKLEVOLKDEGNNNNAEYAREKELEKTTLTEKRAE	327		

Query Match	7.8%;	Score 123.5;	DB 6;	Length. 630;
Best Local Similarity	24.1%;	Pred. No. 0.25;		
Matches	77;	Conservative	51;	Mismatches 101; Indels 91; Gaps 18;
QY	42	MKGVDITHHQF----	YEQVKS--	NPSAOVLNMNTIOKVFQKQ-----YGSGL 83
Db	189	LKG--ITFERIALKSELETYKNSGLSPTSSELAALTKV-	KSLKKEKREBELQPSGNKSKLL	245
QY	84	DD-----	KVDDITAEKKKQYGVYQV----	LSQAGMTLETRKAQITPSK 125
Db	246	EDYIOKHSIDISEKIKALTDDELKRETKQFPDPSKKRL	TELENDLSTFKKELETETKTO--TSK 303	
QY	126	IIVL-----	AVKVAEAEI--	TDAVYKARDETPDYTAIIILNN--EDKAK---- 169
Db	304	KFNLEERKDKEIVKINKLELELLKADKNSGAKKELLEKVS	KSELEITLSELEDDKSVYKQ 363	
QY	170	--EVLKAKAGADPAOLAKDNST--	DEKTKENGEGIT--	FDSASTEVPQVYKKA 218
Db	364	HDELKEQYKREKNGEOLQVTKDYSTTKKIDELQKELD	ALDSPKDKETISAKLTOSTS-- 421	
QY	219	AFALDDVGVSVYITATGTAQVSSOYIVKLTJK--	TEKSSNIDYKREKIKTAVIATQKONDS 277	
Db	422	----DLEAKNKILLILISEKEKTEQOELEKTKHAELEK	LID-----KNKDS 465	
QY	278	TEVOSIIGKELOAAINIKYKD 297		
Db	466	SKTNS----DLEKANKKVLD 481		

RESULT 9
 US-10-299-636-96
 Sequence 96, Application US/10299636
 GENERAL INFORMATION:
 APPLICANT: Billes, David E
 APPLICANT: McDaniel, Larry S
 APPLICANT: Swatilo, Edwin
 APPLICANT: Yoether, Janet
 APPLICANT: Crain, Marilyn J
 APPLICANT: Hollingshead, Susan
 APPLICANT: Tart, Rebecca
 APPLICANT: Brooks-Walter, Alexis
 TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
 FILE REFERENCE: 57909/361
 CURRENT APPLICATION NUMBER: US/10/299,636
 CURRENT FILING DATE: 2002-11-19
 PRIOR APPLICATION NUMBER: 08/714,741
 PRIOR FILING DATE: 1996-09-16
 PRIOR APPLICATION NUMBER: 08/529,055
 PRIOR FILING DATE: 1995-09-15
 NUMBER OF SEQ ID NOS: 111
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 96

LENGTH: 588
 TYPE: PRF
 ORGANISM: Streptococcus pneumoniae
 US-10-299-636-96

Query Match 7.7%; Score 123; DB 6; Length 588;
 Best Local Similarity 25.1%; Pred. No. 0.25;

Matches 64; Conservative 44; Mismatches 103; Indels 44; Gaps 11;

```

OY 72 OKVEFKOYGSGLD-----KEVDFIAEKKOYGNRYVLSQAMTLETTRA 119
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 8 OSKAEKDYDAKKKAKNAKKAVEDOKALDDAKAAQKK-YDDQKTEEKAAL---EKA 62
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 120 QIRTSKVELAVKVAEAEELTDEAYKKAFADEYTPDVTAOIIRLNEDKAEVLEKAKAE- 178
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 63 ---ASEEMDKAVAAVQAQVL---AYQAQADKAKAADAQMT-----DEAKKEEEKTKYF 111
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 179 -----GADFAQLAKDNSTDEKTEKENGGEIT--FDSASTVEPYQYKKAFAALVDGVSD 229
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 112 NTVRAVVYEPPEQLAETKKKSEAKOKAPELTKLEAKAKLEAEKKEATKAKQVDAE 171
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 230 VTTATGTQAVSSQYIVKTKTEKSSNIDY--KEKTKVI---LTQKNDSTFYOSTIG 285
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 172 VAPQAKIALELVQVHRLEQELKEIDSESEDYAKGCFRAPLOSKLDKAKAKLSKEESD 231
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 286 K--ELQAAANIKYKQD 298
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 232 KIDELDAELAKLEDQ 246
   | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10
 US-10-603-113-14747

; Sequence 14747, Application US/10603113

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/10/603,113

; CURRENT FILING DATE: 2003-06-24

; PRIOR APPLICATION NUMBER: US/09/248,796

; PRIOR FILING DATE: 1999-02-12

; NUMBER OF SEQ ID NOS: 28206

; SEQ ID NO 14747

; LENGTH: 901

; TYPE: PRF

; ORGANISM: Candida albicans

US-10-603-113-14747

Query Match 7.7%; Score 122; DB 6; Length 901;
 Best Local Similarity 24.0%; Pred. No. 0.49;

Matches 82; Conservative 40; Mismatches 135; Indels 84; Gaps 15;

```

OY 5 LKGVDMKKLLAGATITLTSVATLAACSKSGADLISMGDVITTEHOFYQVKSNSAQO 64
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 532 LKEVDETEKIPVEE-----SKKVDNDNDDETDLKESKETK---VEKASDPITHE 576
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 VLLNNTIOKVEFEKOGSELDKEVDDTIAEKKOYGE--NYQRY---LSQAMTLETR-K 118
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 577 ELIDEDIEKA--KHOSNEKDDANT--DVLAEGBALVNEDDYDLVKKELKEEGITTEEAP 633
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 119 AQIRTSK-----LVELAVKVAEAEITD---EAYKKAFADEYTPDVTAOIIRLNED 166
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 634 LNIKEKSEKTVQGEVDLVSSVKKDEKPILDVESKEAKKAADDE-----KASQ 684
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 167 KKEVLEKAKAGADFAQLAKDNSTDEKTEK-----GGETTDSAST 209
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 685 PPLDIVESSKE-----TEEKETDETTELPGVTKEVDESEKDTTLLKSEVELEKSE 737
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 210 EYPEOVKKAFAALVDGVSDVITATGTQAVSSQYIVKTKTEKSSNIDYKREKLKTYI 269
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 738 EGPDLIKKEVETATDVATEKSKDVEQAVSS-----TTKETTYEVLVETEKPK-PAYA 790
   | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY 270 LTQKNDSTFYOSTIGKEILOAANIKYKQDAFQNIIFYICG 310
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 791 DDDDDLDLDDISPEETRKHLESQPV-----YIFTSIAGG 823
   | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 11

US-10-273-573-9033

; Sequence 9033, Application US/10273573

; GENERAL INFORMATION:

; APPLICANT: Hysq, Inc

; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 21272-066

; CURRENT APPLICATION NUMBER: US/10/273,573

; CURRENT FILING DATE: 2002-10-18

; PRIOR APPLICATION NUMBER: 09/522,929

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: 09/770,160

; PRIOR FILING DATE: 2001-01-26

; NUMBER OF SEQ ID NOS: 10994

; SOFTWARE: Custom

; SEQ ID NO 9033

; LENGTH: 802

; TYPE: PRF

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (653)..(707)

; OTHER INFORMATION: kinesin light chain repeat proteins domain identified by

; OTHER INFORMATION: EMATRIX, accession number B101160B, p-value=5.424e-09, raw sco

; US-10-273-573-9033

Query Match 7.7%; Score 121.5; DB 6; Length 802;
 Best Local Similarity 20.4%; Pred. No. 0.46;

Matches 60; Conservative 66; Mismatches 107; Indels 61; Gaps 11;

```

OY 35 ESADLISMKGDVITTEHOFYQVKSNSAQOVLNMTIOKVEFKOYGSGLDKEVDVTIAE 94
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 431 EIGELIPLRQKVAEORRTAQOLRAEKASVAFQLSML-----KKAGILAE 475
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 95 EKKQY-----ENYQVLSQAGTITLTRKAQIR-----TSKVELAV 131
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 476 ENRGIGERANLGRQFLVELDQAREKYVQELAAVRADATETRLAEVQREARQSTARELEVNT 535
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 132 KRYVAEELTDEAYKKAFADEYTPDVTAOIIRLN--NEDKAEVLEKAKAGADFAQLAKDN 189
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 536 AKYEGAKVAVLEERQFOEEROKLTAQVQLVFPQREOTKQYBELSK-KLADSDQASKVQ 594
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 190 STDEK-TKENGGEITFDSASTVEPYQYKKAFAALVDGVSDVITATGTQAVSSQYIVKL 248
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 595 OQKLKAVQAQGE-----SQGEAQRLAQ-----LNEIQALQSKQEAHVKLQM 640
   | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 249 TK-KTE---KSSNIDYKREKLKTVILITQKNDSTFYQS--ITKEILOAANIKYK 297
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 641 ERAKTHYDAKKQOONELQOLRLTLEQOLKEKRELRAEERLQHELOAQALKTKE 694
   | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12

US-10-273-573-9040

; Sequence 9040, Application US/10273573

; GENERAL INFORMATION:

; APPLICANT: Hysq, Inc

; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 21272-066

; CURRENT APPLICATION NUMBER: US/10/273,573

; CURRENT FILING DATE: 2002-10-18

; PRIOR APPLICATION NUMBER: 09/522,929

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: 09/770,160

; PRIOR FILING DATE: 2001-01-26

; NUMBER OF SEQ ID NOS: 10994

; SOFTWARE: Custom

; SEQ ID NO 9040

```

; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (169)..(217)
; OTHER INFORMATION: Intermediate filaments protein domain identified by eMatrix,
; OTHER INFORMATION: accession number BL00226b, p-value=3.093e-09, raw score of 23.86
; US-10-293-573-9040

```

Query Match 7.7%; Score 121.5; DB 6; Length 1154;
 Best Local Similarity 20.4%; Pred. No. 0.73; Indels 61; Gaps 11;

Matches 60; Conservative 66; Mismatches 107; Indels 61; Gaps 11;

```

OY 35 EGADLISMKGDTYTHQFYEQVSNPSAQVLLNMTIQKYEKQYSGELDKDEYDTIAE 94
DB 796 ELGLILPRLKRVKAEQERTAOQLRAEKASVAEQLSML-----KKAHGLIAE 840
OY 95 EKQYQ-----ENQYVLSQAGMTLETRKAQIR-----TSKLVELAV 131
DB 841 ENRGLGERALDGRQFLEVLDAQREKIVQELAAVRADAETRLAEVOREAOSTARELEVMT 900
OY 132 KVAEAEITDEAYKKAPEDEYTPDTAQIIRLN--NEDKAEVLEKAKAGADPAOLAKON 189
DB 901 AKYEGAKVYLERQRFQERQKLTQAQVEQLEFQHQRTQVLELSK-KLADSDQSKVQ 959
OY 190 STDEK-TKENGGETPDSASTVEPEQYKKAFAALDVGVSDVITANGTOAYSSQYIVKL 248
DB 960 QOKLKAVQAGGE-----SQEAQRLOAQ-----INELQALQSKQEAHEHYKLM 1005
OY 249 TK-KTE---KSSNIDYKERTKIVITLQKONDSFVOS--IIGKELOANIKYVD 297
DB 1006 ERAKTHYDAKKQONQELQELRLLEQLQENKELRAERLGHLEQDAGLTKTE 1059

```

RESULT 13

US-10-299-636-103

; Sequence 103, Application US/10299636

; GENERAL INFORMATION:

; APPLICANT: Briles, David E

; APPLICANT: McDaniel, Larry S

; APPLICANT: Swiatlo, Edwin

; APPLICANT: Yotner, Janet

; APPLICANT: Crain, Marilyn J

; APPLICANT: Hollingshead, Susan

; APPLICANT: Tarr, Rebecca

; APPLICANT: Brooks-Walter, Alexis

; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF

; FILE REFERENCE: 57909/361

; CURRENT APPLICATION NUMBER: US/10/299,636

; PRIOR FILING DATE: 2002-11-19

; PRIOR APPLICATION NUMBER: 08/714,741

; PRIOR FILING DATE: 1996-09-16

; PRIOR APPLICATION NUMBER: 08/529,055

; NUMBER OF SEQ ID NOS: 111

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 103

; LENGTH: 336

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-10-299-636-103

Query Match 7.5%; Score 119; DB 6; Length 336;
 Best Local Similarity 21.9%; Pred. No. 0.24;

Matches 70; Conservative 45; Mismatches 116; Indels 88; Gaps 11;

```

OY 4 YLKGVDMMKLLAGITLITLVAITLAAKSGEGADLISMKGDTYTHQFYEQVSNPSAQ 63
DB 27 HTQVNLNLIK--SAIKTKYLYELSVLKENSKELSLKTAELT--AAEQFKKDLKP 82
OY 64 QVLLNMTIQKYEKQYSGELDKDEYDTIAEKQYGENYQVLSQAGMTLETRKAQIR 123

```

```

DB 83 E-----KVAEAEKKEVEAKKAKDQKEEDRNPTN-----TYKLELEIAE 125
OY 124 SKLVELAVKRVKAEITDEAYKKAPEDEYTPDTAQIIRLNEDKAEVLEKAKAGADPA 183
DB 126 SDV-----KVRKAEI--ELVKEANE-----SRNEKIKQAEKESKAEAT 166
OY 184 QLKONDSDEKTEKENGGETPDSASTVEPEQYKKAFAALDVGVSDVITANGTOAYSSQY 243
DB 167 RLK-IKTDKKAE-----BEAKRAEESSEKKAFAASKLDTKR----- 203
OY 244 YIVKLTKTEKSSNIDYKERTKIVITLQKONDSFVOSIIGKELOANIKYKQAFQNI 303
DB 204 --AKLSLELSQKIDE-----LAEIYAKLEVQLKDAEGNNV 238
OY 304 FQYIGGGSSSSSSSTNE 322
DB 239 VEAYFEKLEKTYAEKKA 257

```

RESULT 14

US-10-299-636-105

; Sequence 105, Application US/10299636

; GENERAL INFORMATION:

; APPLICANT: Briles, David E

; APPLICANT: McDaniel, Larry S

; APPLICANT: Swiatlo, Edwin

; APPLICANT: Yotner, Janet

; APPLICANT: Crain, Marilyn J

; APPLICANT: Hollingshead, Susan

; APPLICANT: Tarr, Rebecca

; APPLICANT: Brooks-Walter, Alexis

; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF

; FILE REFERENCE: 57909/361

; CURRENT APPLICATION NUMBER: US/10/299,636

; PRIOR FILING DATE: 2002-11-19

; PRIOR APPLICATION NUMBER: 08/714,741

; PRIOR FILING DATE: 1996-09-16

; PRIOR APPLICATION NUMBER: 08/529,055

; NUMBER OF SEQ ID NOS: 111

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 105

; LENGTH: 354

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-10-299-636-105

Query Match 7.5%; Score 119; DB 6; Length 354;
 Best Local Similarity 25.8%; Pred. No. 0.25;

Matches 59; Conservative 43; Mismatches 95; Indels 32; Gaps 10;

```

OY 86 KEVDITAEERKQYGENYQVLSQAGMTLETRKAQIRTSKLVELAVKVAEAEITDEAYK 145
DB 17 KALDAAKAAQK-YDEQOKTEEKAL-----EKA--ASEEMKAAVAQAAVL--AQ 65
OY 146 KAFDEYTPDTAQIIRLNEDKAEVLEKAE-----GADFAQLAKNDSTDEKTE 197
DB 66 QATDKRAADADAKKI-----DEAKKREEAKTFKFNVRAMVVEPDLAETKKKSEANQ 120
OY 198 NGGEIT--PDSASTVEPEQYKKAFAALDVGVSDVITANGTOAYSSQYIYVLTKEKTS 255
DB 121 KAPELTKLEAKAEKAEKKAETAKQYDAEVAAPQAKIAELQVHRLDELKEIDE 180
OY 256 SNIDY-KERLKTIVY--LTKONDSFVOSIIGK--ELOANIKYKQD 298
DB 181 SESEDAKKEGFRAPLOSILDAKAKAKSLKLELSDKIDELDAETAKLEDQ 229

```

RESULT 15

US-10-257-377-3

; Sequence 3, Application US/10257377

; GENERAL INFORMATION:

; APPLICANT: B. Michael R. Woloski,

```

; APPLICANT: Ashley M. Williams,
; APPLICANT: Terrance J. Sereida,
; APPLICANT: Deanna J. Wiebe
; TITLE OF INVENTION: RHAM Peptide Conjugates
; FILE REFERENCE: BERE-100105
; CURRENT APPLICATION NUMBER: US/10/257,377
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/198,613
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-257-377-3

```

```

Query Match      7.5%; Score 119; DB 6; Length 713;
Best Local Similarity 20.4%; Pred. No. 0.6;
Matches 78; Conservative 69; Mismatches 126; Indels 110; Gaps 16;

```

```

QY 31 SKGSEADLISMKGVIT-----EHQFYQVKSNPQAQVYL----- 67
DB 305 SKDREAEESLSAEMOYLTEKLLERQVEYKQLQONELQSOSLLQOEKLSAHLQOQLCSFQ 364
QY 68 -NNTIOK-VFEKQYGSGLDDKEVDPTIAEKKQYGENYQVLSQAGMTLE--TRKAQIR 123
DB 365 EEMTSERNVFKQDLALDEL---DAVQQRKEQSEKLYKQLEETKSTADQLRLDDLR 421
QY 124 SKLVLAVKVAEAEITDEAYRKAFD-EYTPDVTAI-----IRLNN---- 164
DB 422 EKEIELEKRTAAHAQATVIAQEKYSDPTAQLRDVTAQLSEYKSSSTLKEIEDLKLENTLQ 481
QY 165 -----EDKAKVLEKAKAEGADFAQLAKD--NS-----TD- 192
DB 482 EKVAAAEKRVEDVQOQIL-TAESTNQEVAKVQODLQNSSTLKEAIEIKETSYLERITDL 540
QY 193 -----EKTRENGGEIT-FDSASTVEPEQVKK-----AALPDVGV 227
DB 541 ONQLRQONEDFRKQLEPEGAQKTEKETATVETLMEINKRRLLYEELPDKTKPFQOQDDAF 600
QY 228 SDVITATGTQAYSSQYIVKLRKTEKSSNIDDKYKELKTVILITOKONDSTFVQSIIGK- 286
DB 601 EAEKQALLNEHGATQPOLSKIRDSYAQLGHQNLKQKIKHVYKLDKDNSQ--LKSEVSKL 658
QY 287 ELQANIKYKQDAFQNIPTQYIG 309
DB 659 RSQLAARKQNELRLQGEIDKALG 681

```

Search completed: August 29, 2003, 10:44:31
Job time : 26 secs

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:19:19 ; Search time 83 Seconds

(Without alignments)
615.783 Million cell updates/sec

Title:

US-10-049-473a-2

Perfect score:

1588

Sequence:

1 SNTYLGKGVDMKKILAGAIR.....IFQYIGGDSSTSSSTSSNE 322

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

1107863 seqs, 158726573 residues

1107863

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1588	100.0	322	22	AA873245
2	1541	97.0	313	24	ABU01392
3	1541	97.0	314	21	AAV81675
4	916	57.7	213	19	AAW5079
5	916	57.7	213	23	ABP54573
6	736	46.3	309	23	ABP30803
7	736	46.3	309	23	ABP30888
8	717	45.2	298	23	ABP30081
9	696	43.8	351	23	ABP28729

10	625	39.4	317	23	ABP30009	Streptococcus poly
11	605.5	38.1	342	20	AAV00100	Enterococcus faeca
12	605.5	38.1	342	20	ABP43319	E faecalis EP049 p
13	605.5	38.1	342	24	ABU13598	Enterococcus faeca
14	562.5	35.4	322	20	AAV00101	Enterococcus faeca
15	562.5	35.4	322	23	ABP43320	E faecalis EP049 a
16	562.5	35.4	322	24	ABU13599	Enterococcus faeca
17	516	32.5	217	23	ABP28728	Streptococcus poly
18	511	32.2	308	23	ABW55074	Lactococcus lactis
19	393.5	24.8	293	23	ABW94437	Listeria monocytog
20	375.5	23.6	299	22	AAV3246	Protease maturatio
21	374	23.6	299	22	AAV3248	Protease maturatio
22	374	23.6	299	22	AAV3248	Protease maturatio
23	328	20.7	294	23	ABW47961	Listeria monocytog
24	322	20.3	74	19	AAV38711	Streptococcus pneu
25	212.5	13.4	320	24	ABU18994	Pathogen specific
26	210	13.2	330	23	ABP38966	Staphylococcus epl
27	204	12.8	320	18	AAW89744	Staphylococcus aur
28	173	10.9	42	18	AAV11343	S. pneumoniae prot
29	155.5	9.8	336	20	AAV19969	B. burgdorferi ant
30	153.5	9.7	317	20	AAV19969	B. burgdorferi ant
31	136.5	8.6	684	16	AAV33912	Streptococcus pneu
32	135.5	8.5	619	15	AAV33437	Pneumococcal surfa
33	135.5	8.5	619	17	AAV33437	Pneumococcal surfa
34	135.5	8.5	619	20	AAV41838	Streptococcus pneu
35	135.5	8.5	619	23	AAE18782	S. pneumoniae Rxi
36	135.5	8.5	619	19	AAV70336	Pneumococcal surfa
37	135.5	8.5	648	19	AAW62274	Streptococcus pneu
38	135.5	8.5	648	20	AAV41837	Streptococcus pneu
39	135.5	8.5	648	20	AAW87879	A. pneumoniae trun
40	135.5	8.5	653	20	AAW92456	Pneumococcal surfa
41	130.5	8.2	619	17	AAW86911	Streptococcus pneu
42	130.5	8.2	2478	22	AAU34370	Staphylococcus aur
43	130.5	8.2	2478	22	AAU34370	Staphylococcus aur
44	130.5	8.2	2478	24	ABU19002	Pathogen specific
45	130	8.2	288	14	AAV35436	B. burgdorferi str

ALIGNMENTS

RESULT 1	AA873245	standard; Protein: 322 AA.
ID	AA873245;	
XX	AC	AA873245;
XX	DT	14-MAY-2001 (first entry)
XX	DE	Protease maturation protein, Pmp.
XX	KW	Antibacterial; protease maturation protein; Pmp; bacterial infection;
XX	KW	vacuole.
XX	OS	Streptococcus pneumoniae.
XX	PN	EP1075841-AA
XX	PD	14-FEB-2001.
XX	PT	13-AUG-1999; 99EP-0202640.
XX	PR	13-AUG-1999; 99EP-0202640.
XX	PA	(UYRO-) UNIV ROTTERDAM ERASMUS.
XX	PI	De Groot R, Hermans PMW;
XX	DR	WPI: 2001-236756/25.
XX	DR	N-PSDB: AAF75748.
XX	PT	New vaccine comprising a protease maturation protein from Streptococcus pneumoniae useful for treating or preventing microbial infections.

PT specifically streptococcal infections -

XX Claim 4; Fig 1; 18pp; English.

PS Streptococcus pneumoniae is a major cause of invasive diseases such as

CC meningitis, bacteraemia and pneumonia, as well as non-invasive diseases

CC such as acute otitis media and sinusitis. The present invention relates

CC to a vaccine comprising Protease Maturation Protein (Pmp) of

CC S. pneumoniae (the present sequence). The vaccine can be used for the

CC treatment of S. pneumoniae infections. Pmp is a surface-associated

CC protein.

CC

XX Sequence 322 AA;

SO

Query Match 100.0%; Score 1588; DB 22; Length 322;

Best Local Similarity 100.0%; Pred. No. 2,4e-113;

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SNTVLKGVDMKKKLLAGATITLLSVATLAACSGSGADLISMKGDVTEHQFYEQVKNP 60

DB 1 SNTVLKGVDMKKKLLAGATITLLSVATLAACSGSGADLISMKGDVTEHQFYEQVKNP 60

QY 61 SAQGVLLNMTIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQ 120

DB 61 SAQGVLLNMTIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQ 120

QY 121 IRTSKLVEIAVKKAAEELDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKEGA 180

DB 121 IRTSKLVEIAVKKAAEELDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKEGA 180

QY 181 DPAQLANDNSTDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYS 240

DB 181 DPAQLANDNSTDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYS 240

QY 241 SQYIYVLTTKTEKSSNIDYKEKLTAVILTKQONDSFVOSITIGKELQANIKVKDAQF 300

DB 241 SQYIYVLTTKTEKSSNIDYKEKLTAVILTKQONDSFVOSITIGKELQANIKVKDAQF 300

QY 301 QNIFTQYIIGGSSSSSSSTSN 322

DB 301 QNIFTQYIIGGSSSSSSSTSN 322

RESULT 2

ABU01392

ID ABU01392 standard; Protein; 313 AA.

XX

AC ABU01392;

DT 11-FEB-2003 (first entry)

XX

DE 5. pneumoniae type 4 strain protein from coding region #967.

XX

KM Bacterial meningitis; pneumonia; sepsis; otitis media;

KM ear infection; antiinflammatory; antibacterial; immunostimulant;

KM auditory; respiratory; gene therapy; vaccine.

XX

OS Streptococcus pneumoniae type 4 strain.

XX

PN MO200277021-A2.

XX

PD 01-OCT-2002.

XX

PF 27-MAR-2002; 2002MO-IB02163.

XX

PR 27-MAR-2001; 2001GB-0007658.

XX

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

DR MPI; 2003-040579/03.

DR N-PSDB; ABX06679.

XX

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,

PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

PT or ear infection

PS

XX Claim 1; SEQ ID No 1934; 56pp; English.

CC The invention relates to a protein comprising or having at least 50%

CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,

CC expressed from 2469 of 2489 identified DNA coding regions from the

CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as

CC AB556454. Also included are an antibody which binds one of the

CC proteins, treating a patient by administering the protein, DNA or

CC antibody (in a composition), a kit comprising first and second primers,

CC which are the nucleic acid cited above or fragments between nucleotides

CC 8-100 of a sequence not defined in the specification, for amplifying a

CC target sequence contained within a Streptococcus nucleic acid sequence,

CC where the first primer is substantially complementary to the target

CC sequence and the second primer is substantially complementary to the

CC complement of the target sequence, and where the parts of the primers

CC having substantial complementarity define the termini of the target

CC sequence to be amplified, assay comprising contacting a test compound

CC with the protein, and determining whether the test compound binds to the

CC protein and a Streptococcus pneumoniae bacterium, where one or more

CC genes encoding the proteins has been rendered inactive. The proteins,

CC nucleic acid molecules, antibody and compositions are useful as

CC medicaments for treating or preventing a disease or infection due to

CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,

CC sepsis, otitis media or ear infection. They are also useful in developing

CC vaccines, diagnostics and antibiotics. The methods are useful for

CC identifying immunodominant proteins. The present sequence is one of

CC the 2469 proteins expressed by the identified coding regions from the

CC genomic sequence.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

SO Sequence 313 AA;

Query Match 97.0%; Score 1541; DB 24; Length 313;

Best Local Similarity 100.0%; Pred. No. 9e-110;

Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVTEHQFYEQVKNPSPAQVLLNM 69

DB 1 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVTEHQFYEQVKNPSPAQVLLNM 60

QY 70 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVEL 129

DB 61 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKVEL 120

QY 130 AVKKAFAELTPEAVKKAAPDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 189

DB 121 AVKKAFAELTPEAVKKAAPDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 180

QY 190 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 249

DB 181 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 240

QY 250 KTEKSSNIDYKELKTYIILTKQONDSFVOSITIGKELQANIKVKDAQFONIFTQYIG 309

DB 241 KTEKSSNIDYKELKTYIILTKQONDSFVOSITIGKELQANIKVKDAQFONIFTQYIG 300

QY 310 GGDSSSSSSSTSN 322

DB 301 GGDSSSSSSSTSN 313

RESULT 3

ID	AA	AA81675 standard; Protein; 314 AA.
XX	AA81675;	
XX	24-MAY-2000	(first entry)
XX	Streptococcus pneumoniae	protein sequence ID212 - 4136.1.
XX	Streptococcus pneumoniae; vaccine; screening; protein antigen;	
XX	antibacterial; antiinflammatory; meningitis; infection; diagnosis;	
XX	pneumococcal disease.	
XX	Streptococcus pneumoniae.	
XX	WP030006774.2.	
XX	10-FEB-2000.	
XX	27-JUL-1999;	99WO-GB02451.
XX	27-JUL-1998;	98GB-0016337.
XX	19-MAR-1999;	99US-0125164.
XX	(MICR-) MICROBIAL TECHNIQS LTD.	
XX	Gilbert CFG, Hansbro PM;	
XX	WPI: 2000-195300/17.	
XX	New Streptococcal protein, useful as a vaccine, for diagnosis of	
XX	pneumococcal diseases and for screening agents capable of antagonizing	
XX	or inhibiting expression of the protein	
XX	Claim 6; Page 92; 108pp; English.	
XX	AA81501 to AA81679 represent specifically claimed protein sequences	
XX	isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent	
XX	specifically claimed nucleotide sequences isolated from S. pneumoniae.	
XX	The sequences have antibacterial and antiinflammatory properties.	
XX	The protein sequences, and fragments of them, are useful as immunogens	
XX	and/or antigens. The nucleotide sequences can be used in vaccines and in	
XX	diagnostic assays. The proteins and nucleotides can be useful for the	
XX	detection and diagnosis of S. pneumoniae. The protein sequences are also	
XX	useful for screening an agent capable of antagonizing, inhibiting or	
XX	interfering with the function or expression of the proteins in which the	
XX	agent is useful for treatment or prophylaxis of S. pneumoniae infection	
XX	and meningitis. AA05591 to AA05614 represent primers used in the	
XX	exemplification of the present invention.	
XX	Sequence 314 AA;	
XX	Query Match 97.0%; Score 1541; DB 21; Length 314;	
XX	Best Local Similarity 100.0%; Pred. No. 9e-110;	
XX	Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
XX	10 MKKTLIAGATITLVAATLAAGSGSEGADLSMKDVTITEHOFEQVSNPSAQOVLNM 69	
XX	1 MKKTLIAGATITLVAATLAAGSGSEGADLSMKDVTITEHOFEQVSNPSAQOVLNM 60	
XX	70 TIOKFEKYOYGESELDKEVDOTIAEKKYOYGENYORVLSQAGMTLETTRKAQRTSKIVEL 129	
XX	61 TIOKFEKYOYGESELDKEVDOTIAEKKYOYGENYORVLSQAGMTLETTRKAQRTSKIVEL 120	
XX	130 AVKKAFAELTDEAVKKAFADETPPVTAQIITLNEDNKAKEYLEKAKAGAPAOALAKDN 189	
XX	121 AVKKAFAELTDEAVKKAFADETPPVTAQIITLNEDNKAKEYLEKAKAGAPAOALAKDN 180	
XX	190 STDEKTEKNGEITPDSASTEVEQYKKAFAFDVGVSDVITATGTQAYSSQYYIVKLT 249	
XX	181 STDEKTEKNGEITPDSASTEVEQYKKAFAFDVGVSDVITATGTQAYSSQYYIVKLT 240	
XX	230 KTEKSSINDYKEKLTQVITLTKQNDSTFVOSITIGKELOAANIKVQDAFONITFYIG 309	

Db	241 KKFKESSSIDDKKELKTVLLTQGNDSFPVQSITGKELQAANIKVKQAQNIFTYIG 300
QY	310 GDDSSSSSTSN 322
Db	301 GDDSSSSSTSN 313
RESULT 4	
AAM55079	ID AAM55079 standard; Protein; 213 AA.
AC	AAM55079;
XX	02-OCT-1998 (first entry)
Dt	Streptococcus pneumoniae SP021 protein.
DE	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW	detection; pneumonia; otitis media; meningitis.
KX	Streptococcus pneumoniae.
OS	MO9818930-A2.
PN	07-MAY-1998.
PD	30-OCT-1997; 97WO-US19422.
PF	31-OCT-1996; 96US-0029960.
PR	(HUMA-) HUMAN GENOME SCI INC.
PA	Choi GH, Hromocky J A, Johnson LS, Kunsch CA;
PI	WP1: 1998-272224/24.
DR	N-PSDB: AAV27340.
XX	
Pt	Nucleic acid encoding antigenic peptide(s) from Streptococcus
Pt	pneumonia - or their epitope-containing fragments, useful in
Pr	protective or therapeutic vaccines, and for diagnosis
XX	
Ps	Claim 11; Page 55; 118bp; English.
CC	The present sequence represents a protein from Streptococcus pneumoniae.
CC	The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC	can be useful in vaccines for inducing protective antibodies against
CC	Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC	pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC	are used to detect Streptococcus infection (by usual hybridisation or
CC	amplification methods), also for isolating Streptococcus genes or their
CC	allelic variants. The protein can be used similarly to detect specific
CC	antibodies in standard immunoassays, especially for diagnosing or
CC	monitoring infections. Antibodies which bind the protein are used to
CC	detect corresponding antigens, to purify the protein and for passive
CC	immunisation (optionally coupled to a toxin). Vaccines are administered,
CC	e.g. by injection, orally or through the skin, typically at 0.01-1000
CC	(especially 10-300) mu g/ml per dose.
CC	
XX	
SQ	Sequence 213 AA:
Query Match	57.7%; Score 916; DB 19; Length 213;
Best Local Similarity	93.6%; Pred. No. 3..3e-62;
Matches 190; Conservative	0; Mismatches 5; Indels 8; Gaps 2
QY	31 SKSGSAGDLISKGVITEHQFYEQYKSNPAAOVLNMTTOKVEEKOYGSLDKEVD 90
Db	1 SKSGSAGALISKKGVITEHQFYEQYKSNPAAOVLNMTTOKVEEKOYGSELDPKEVD 60
QY	91 TTAEEKKQGYENYOVLSDAGMTLETRAKQIRTSKLVELAVKVAEAEITDAEYKARDE 150
Db	61 TTAEEKKQGYENYOVLSDAGMTLETRAKQIRTSKLVELAVKVAEAEITDAEYKARDE 120

```

OY 151 YTPDVTAGIIRLNEDKAEVLEKAKAEGADFAOLADNSTDTEKNGEITFDASASTE 210
DB 121 YTPDVTAGIIRLNEDKAEVLEKAKAEGADFAOLADNSTDTEKNGEITFDASASTE 180
OY 211 VP-EQVKKAFA-----LDVD 225
DB 181 VPGASPKKPLFAFRGCMVFLDLD 203

RESULT 5
ABP54573
ID ABP54573 standard; Protein; 213 AA.
AC ABP54573;
XX
XX
XX 04-SEP-2002 (first entry)
DE
XX
XX
XX S. pneumoniae SP021 protein sequence SEQ ID NO:34.
XX
XX Streptococcus pneumoniae; epitope: vaccine; antigenic protein;
XX Streptococcus pneumoniae; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX 22-JAN-2001; 2001US-0765272.
XX
XX 30-OCT-1997; 97US-0961083.
XX
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX WPI: 2002-479261/51.
XX N-PSDB; AB084808.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
XX Streptococcus and for preventing or attenuating disease caused by
XX Streptococcus infection.
XX
XX Claim 11; Page 24; 70pp; English.
XX
XX AB084792 to AB084904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54573 to ABP54669.
XX The S. pneumoniae antigens have antibacterial activity and can be
XX used in vaccines. The S. pneumoniae antigens can also be used to
XX prevent or attenuate a Streptococcal infection in an animal. The
XX polynucleotides encoding the S. pneumoniae antigens can be used to
XX detect Streptococcus nucleic acids. AB084905 to AB085130 represent
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX which are used in an example from the present invention.
XX
XX
XX Sequence 213 AA;
SQ
Query Match 57.7%; Score 916; DB 23; Length 213;
Best Local Similarity 93.6%; Pred. No. 3.3e-62;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;
OY 31 SKGSEGADLISMGDVTTEHOFYEQYKSNPSAQOVLNNTIOKVFKEQYSGELDDKEVND 90
DB 1 SKGSEGADLISMGDVTTEHOFYEQYKSNPSAQOVLNNTIOKVFKEQYSGELDDKEVND 60
OY 91 TIAEKKQYGENYORVLSQAGMTLETRKAQIRTSKIVELAVKVAEALITDEAYKKAFFE 150

```

```

DB 61 TIAEKKQYGENYORVLSQAGMTLETRKAQIRTSKIVELAVKVAEALITDEAYKKAFFE 120
OY 151 YTPDVTAGIIRLNEDKAEVLEKAKAEGADFAOLADNSTDTEKNGEITFDASASTE 210
DB 121 YTPDVTAGIIRLNEDKAEVLEKAKAEGADFAOLADNSTDTEKNGEITFDASASTE 180
OY 211 VP-EQVKKAFA-----LDVD 225
DB 181 VPGASPKKPLFAFRGCMVFLDLD 203

RESULT 6
ABP30803
ID ABP30803 standard; Protein; 309 AA.
AC ABP30803;
XX
XX
XX 02-JUL-2002 (first entry)
DE
XX
XX Streptococcus polypeptide SEQ ID NO 10782.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX W0200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001MO-GH04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX WPI: 2002-352536/38.
XX N-PSDB; ABN71434.
XX
XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein.
XX
XX Claim 1; Page 4184; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 309 AA;
SQ

```

RESULT 7
 ABP30888
 ID ABP30888 standard; Protein; 309 AA.
 XX
 AC ABP30888;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 10952.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; Infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN MO200234771-A2.
 XX
 PQ 02-MAY-2002.
 XX
 PE 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;
 PI Teteltn H;
 XX
 DR WPI: 2002-352536/38.
 DR N-PSDB: ABN71519.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 4192; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

Oy		11	KKKLLAGATTLLTSVATTLAACS-KGSESGADLISMKGDVTTEHQFYEOGVSNPSAQOVLMM	69
Db		4	RSKLAGELTLLMSVATTLAACSGKSTNGTWVMKGDTITVSDFYDQVKTSKAAQOSMLTL	63
Oy		70	TIOKFVEKYOGSELDDKEVDPTIAEEKROYGENYOVLISOAGMTLETTRKAQIRTSKIYEL	129
Db		64	ILSRFPDYQIGKVSDCKVSEAYNNTAKGYGSPSALSQAGLITEPGYKOQIRTMVLVEY	123
Oy		130	AVKRAVEALETDWAYKKAFADEXTDPVTAQIIRLNNEDKRAKEYLEKAKAGADFAQLAND-	188
Db		124	AVKEAAKKELTFEANYKAEVKNYTPETSYOVQIKLDADDKAKSVLKDVKADGAPFAKIAKEK	183
Oy		189	-NSTDEKREKENGEGTEPDSASTFEVPEQYKKAFFALDDVGVSIVTATGTQAVSSQYYTYK	247
Db		184	TTATDKKY----EKFPDSAGTTLTKREWMAAFKLDKNGVSDVSVSTVSTYKTSYYTIK	238
Oy		248	LTKRKESGNIDYKEKELKTIVTLIQOKONSTFVQSIIQELQANIKVYKDQAFONTFOY	307
Db		239	VTDKREKESDMKSYNRRLKEVILLKDKTSDRAFQNKRVISKALKERANKYIKDKAPAGILSOY	298
Oy		308	ICGGDSSS 315 	
Db		299	ATTSGSSS 306	
RESULT 8				
ABP30081				
ID	ABP30081 standard; Protein; 298 AA.			
XX				
AC	ABP30081;			
XX				
DT				
XX	02-JUL-2002 (first entry)			
DE				
XX	Streptococcus polypeptide SEQ ID NO 9338.			
KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;			
KM	group A streptococcus; Streptococcus pyogenes; antibacterial;			
KX	antlinflammatory; infection; vaccine; meningitis; gene therapy.			
OS				
XX	Streptococcus agalactiae.			
PN				
XX	WO200234771-A2.			
PD	02-MAY-2002.			
XX				
Pf	29-OCT-2001; 2001WO-GB04789.			
XX				
PR	27-OCT-2000; 2000GB-0026333.			
PR	04-NOV-2000; 2000GB-0028727.			
PR	07-MAR-2001; 2001GB-0005640.			
PA	(CHTR-) CHIRON SPA.			

PA (GENO-) INST GENOMIC RES.
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX WPI: 2002-352536/38.
 DR N-PSDB: ABN70712.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1: Page 4058; 4535pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 XX Sequence 298 AA:
 SQ
 Query Match 45.2%; Score 717; DB 23; Length 298;
 Best Local Similarity 48.3%; Pred. No. 8.3e-47;
 Matches 145; Conservative 60; Mismatches 87; Indels 8; Gaps 3;
 OY 19 IITLIVATLAACS-KGSEADLISMKGVITEHOYFQVKSNSAQOVLNMTIQYEEK 77
 Db 1 MTLASVATLAACSGTNSGNTVNTMKDITTVSDYDVQVTSKAAQOSMLTILSRVFD 60
 OY 78 QYSELDKDEVDITIAEKKROYGENTRVLSQAGMTLETAKQIRTSKVELAVKVAEA 137
 Db 61 QYGDVSKDKVSEAYNTAKGSGNSFSSALSGNLTPEGKQOIRTMLEVYAVKEAAK 120
 OY 138 ELTDEAKKAFDEYTPVTAQIIRLNNEDEKAKVELEKAGADFAQLAND--NSIDEKT 195
 Db 121 ELTEANKKEAYKNYTPETSVQVTKLDAEDKAKSVLKADVGADGAFKIAKEKTTATDKV 180
 OY 196 KENGGEITFDASSTEVEQVKKAAFDVGVSDVITATGTQAYSSQYIVYKLTKEKS 255
 Db 181 -----EKKFDSAGTTPKREYMSAFAKLKNGVSDVSTYTKSYITIKYTDTEKK 235
 OY 256 SNIDYKREKTKVILQKONDSTFVOSITIGKELQANIKVQADAFONIFQYIGGSDSS 315
 Db 236 SDMKSYKNRLKEVILKDKTSDRAFONKVISKALEKANVKKIDKAFAGILLISQVATGSSSS 295

RESULT 9

ABP28729 standard: Protein, 351 AA.

AC ABP28729;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 6634.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.
 XX
 XX W0200234771-A2.
 XX
 XX 02-MAY-2002.
 PD
 PD
 XX
 XX 29-OCT-2001; 2001MO-GB04789.
 PE
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX WPI: 2002-352536/38.
 DR N-PSDB: ABN69360.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1: Page 3825; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 XX Sequence 351 AA:
 SQ
 Query Match 43.8%; Score 696; DB 23; Length 351;
 Best Local Similarity 44.7%; Pred. No. 4.2e-45;
 Matches 140; Conservative 62; Mismatches 105; Indels 6; Gaps 2;
 OY 13 KLAGAITLLSVATLAACSGSEADLISMKGVITEHOYFQVKSNSAQOVLNMTIQ 72
 Db 6 KLIASVATLAACSGTNSGNTVNTMKDITTVSDYDVQVTSKAAQOSMLTILSRVFD 60
 OY 73 KYFEKQYSELDKDEVDITIAEKKROYGENTRVLSQAGMTLETAKQIRTSKVELAVK 132
 Db 66 RVEFAQYGDVSKKVEKRAYHNTAEQYGSFSAALQOSLTPETFKRQIRRSKLEVEAVK 125
 OY 133 KYAEAEITDEAKKAFDEYTPVTAQIIRLNNEDEKAKVELEKAGADFAQLANDST 192
 Db 126 EAKKELTTOEKKAKVEESTPTMAVEYITLNEEYAKSVLEKAGADFAIAKEKTT- 184
 OY 193 EKTENGGEITFDASSTEVEQVKKAAFDVGVSDVITATGTQAYSSQYIVYKLTKEKT 252
 Db 185 --TPKKYVYKFDGATNVPDVTYVKAASLNGGJSIDVTSVLDPTSYCKKRYIVKTKKA 242
 OY 253 EKSNIDYKREKTKVILQKONDSTFVOSITIGKELQANIKVQADAFONIFQYIGGCD 312
 Db 243 EKSKWQEKRLKRLKLTIAESKDMFNQKVIANALDKRANVKKIDKAFANILAYANLQ 302
 OY 313 ---SSSSSTSN 322

DB 303 KTKASESSTTSE 315

RESULT 10
ID ABP30009 standard; Protein; 317 AA.
XX AC ABP30009;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ.ID NO 9194.
XX KM Streptococcus: GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX RW antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS Streptococcus pyogenes.

XX PN MO200234771-A2.
PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001MO-GB04789.
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tetzelin H;
XX MPI: 2002-352536/38.
DR N-PSTD: ABN70640.

PT New Streptococcus protein for the treatment or prevention of infection
PR or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX

PS Claim 1; Page 4043; 4525sp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6004-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX

SQ Sequence 317 AA:

Query Match 39.4%; Score 625; DB 23; Length 317;
Best Local Similarity 42.8%; Pred. No. 9.7e-40;
Matches 130; Conservative 62; Mismatches 102; Indels 10; Gaps 4

OY 9 DMK--KLACATILLSVATTAAAGSGEGADLISMKGVDITTEHOFYEQVKNPSAOYL 66
||| :|: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:
DB 8 DMKQNNKITLGVTAVTTAATTSACGSSHNHTKLTAKSMKGDTIVSPDYNETKWTETLAQAM 67
LMNTLOXFKFEKOYGSELDDKEVDPTIAEKKOYGENYORVLSQAQMTLETTRKAQRITSXL 126

[illegible]

CC bacterial genera. The polynucleotides are useful as probes for gene
CC mapping and for identifying E. faecalis in biological samples. Sequences
CC ABU13508-ABU13755 represent EF040 polypeptides of the invention.
CC Note: The sequence data for this patent can also be obtained from USPTO
CC at seqdata.uspto.gov/sequence.html.

XX Sequence 342 AA;

Query Match 38.1%; Score 605.5; DB 24; Length 342;
Best Local Similarity 41.8%; Pred. No. 3.3e-38;
Matches 132; Conservative 62; Mismatches 113; Indels 9; Gaps 5;

DB 10 MKKLLAGATLLSVATLAACSGSEADLISMKGVDITEHQFYEQVKNPSAQOVLNLM 69
1 MKKLLAGATLLSVATLAACSGSEADLISMKGVDITEHQFYEQVKNPSAQOVLNLM 58
OY 70 TIOKVEKQYSELDKEVDITAEKKO--YGENYQVLSQAGMTLETRRAQIRTSKL 126
DB 59 VIKYVEEKKYGDVTKXIOKNFDEKQVEAOGGKFSALQAGLITKTEFKKOLKORAA 118
OY 127 VELAVKVAEELTDEAYKKAPEDEYPTAOIIRLNEDKAKEVLEKAKAGADPAOLA 186
DB 119 YAGLKA--AHLKITDDELTAWASFPEVEAOIIOVASEDDAKAV--KKEITGGDPTKTA 175
OY 187 KDNSTDEKTKENGSEITFEDASTVEPEQYKKAFAALVDVGSVDITATGTOAYSSQYTY 246
DB 176 KEKSDTATKKGCGRIKPFDSQATVPAEYKKAFAALVDVGSVDITATGTOAYSSQYTY 235
OY 247 KLTETKSSNIDDKREKLTLYLTKOKNDSTFVOSIIGKEIQAANIKYKDAFONIFQ 306
DB 236 KMTKNKAKNDMKPEKEIKKIAETKTLADQTFVSKVIDELKAAVNIKIDDAFKNALAG 295
OY 307 YIGGSSSSSSSTPNE 322
DB 296 IY--QTESSASSSEKKE 310

RESULT 14
AAV00101

ID AAV00101 standard; Protein; 322 AA.

XX AAV00101;

DB 20-APR-1999 (first entry)

DE Enterococcus faecalis antigenic polypeptide fragment EF049.

XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

XX detection; attenuation; antigenic.

XX Enterococcus faecalis.

XX WO9850554-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08959.

XX 14-NOV-1997; 97US-0066009.

XX 06-MAY-1997; 97US-0044031.

XX 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Bailey C, Choi GH, Hromockyj A, Kunsch CA;

XX WPI; 1999-070095/06.

XX N-PSDB; AAX20091.

XX New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection

PS Claim 9; Page 127; 301pp; English.

XX The present sequence represents an antigenic polypeptide fragment
CC isolated from Enterococcus faecalis. The present invention describes
CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC The proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the Enterococcus genus in an animal.
CC They can also be used for detecting Enterococcus antibodies in a sample.
CC The nucleotide sequences can be used for detecting Enterococcus nucleic
CC acids. Products from the present invention can also be used for
CC screening compounds to identify agonists and antagonists of E. faecalis
CC protein activity.

XX Sequence 322 AA;

Query Match 35.4%; Score 562.5; DB 20; Length 322;
Best Local Similarity 41.2%; Pred. No. 6e-35;
Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;

DB 30 CSKSGEADLISMKGVDITEHQFYEQVKNPSAQOVLNMTIOKVEKQYSELDKEVD 89
1 CSKSGK--DIATMKGSTIVDDPFYNDIKQSTSQAFSQMVIYKFEKYGKVDKXIQ 58
OY 90 DTIAEKKO--YGENYQVLSQAGMTLETRRAQIRTSKLVELAVKVAEELTDEAYK 146
DB 59 KNEDEKQVEAOGGKFSALQAGLITKTEFKKOLKORAAVDAGLK--AHLKITDDELKT 116
OY 147 AFDEYTPVTAOIIIRLNEDKAKEVLEKAKAGADPAOLAKNSDEKTKENGSEITPDS 206
DB 117 AMASFPEVEAOIIOVASEDDAKAV--KKEITGGDPTKIAKERSDPTATKKGCGRIK 175
OY 207 ASTEVEPEQYKKAFAALVDVGSVDITATGTOAYSSQYTYVKTETKSSNIDDKREKL 266
DB 176 QATVPAEYKKAFAALVDVGSVDITATGTOAYSSQYTYVKTETKSSNIDDKREKL 235
OY 267 TVILTKQNDSTFVOSIIGKEIQAANIKYKDAFONIFQYIGGSSSSSSSTPNE 322
DB 236 KIAETKTLADQTFVSKVIDELKAAVNIKIDDAFKNALAGV--QTESSASSSEKKE 290

RESULT 15
ABP43320

ID ABP43320 standard; Protein; 322 AA.

XX ABP43320;

DB 05-AUG-2002 (first entry)

DE E faecalis EF049 antigenic fragment.

XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

XX Enterococcus faecalis.

XX US2002045737-A1.

XX 18-APR-2002.

XX 04-MAY-1998; 98US-0071035.

XX 04-MAY-1998; 98US-0071035.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2002-425450/45.

XX N-PSDB; ABN98076.

XX New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:32:29 ; Search time 29 seconds
(Without alignments)
469.796 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588
Sequence: 1 SMTYKGVDMKRLLAGAIT.....IFQYIGGDSSTSTSTNE 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata1/1aa/5A.COMB.pep:*
4: /cgn2_6/prodata1/1aa/5B.COMB.pep:*
5: /cgn2_6/prodata1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	57.7	213	3	US-08-961-083-34
2	916	57.7	213	4	US-09-536-784-34
3	634	39.9	344	4	US-09-107-532A-6355
4	605.5	38.1	342	4	US-09-071-035-182
5	588.5	37.1	303	4	US-09-107-532A-6113
6	562.5	35.4	322	4	US-09-071-035-184
7	450.5	28.4	302	4	US-09-107-532A-4736
8	322	20.3	74	4	US-08-858-207A-508
9	210	13.2	330	4	US-09-134-001C-3811
10	147.5	9.3	447	4	US-09-252-991A-17422
11	137.5	8.7	443	4	US-09-328-352-7725
12	135.5	8.5	619	1	US-08-465-746-2
13	135.5	8.5	619	1	US-08-214-164-2
14	135.5	8.5	619	2	US-08-467-852A-3
15	135.5	8.5	619	2	US-08-246-636-2
16	135.5	8.5	619	2	US-08-247-491A-3
17	135.5	8.5	619	2	US-08-319-795-2
18	135.5	8.5	619	2	US-08-468-985-2
19	135.5	8.5	619	3	US-08-312-849-2
20	135.5	8.5	648	1	US-08-072-070-2
21	135.5	8.5	648	1	US-08-469-434-2
22	135.5	8.5	648	1	US-08-214-222-2
23	135.5	8.5	648	2	US-08-467-852A-2
24	135.5	8.5	648	2	US-08-468-718-2
25	135.5	8.5	648	2	US-08-247-491A-2
26	135.5	8.5	648	3	US-08-446-201-3
27	135.5	8.5	695	1	US-08-127-499A-23

28	135.5	8.5	695	1	US-08-482-847-23	Sequence 23, Appl
29	131.5	8.3	588	4	US-08-714-741-42	Sequence 42, Appl
30	131.5	8.3	1231	4	US-08-714-741-41	Sequence 41, Appl
31	129.5	8.2	864	4	US-08-714-741-40	Sequence 40, Appl
32	126.5	8.0	8991	4	US-08-714-741-32	Sequence 32, Appl
33	124	7.8	370	4	US-09-252-991A-29473	Sequence 29473, A
34	124	7.8	485	4	US-09-071-035-34	Sequence 34, Appl
35	123.5	7.8	274	1	US-08-137-175A-8	Sequence 8, Appl
36	123.5	7.8	274	3	US-08-479-017-8	Sequence 8, Appl
37	123.5	7.8	289	1	US-08-072-070-4	Sequence 4, Appl
38	123.5	7.8	289	1	US-08-469-434-4	Sequence 4, Appl
39	123.5	7.8	289	1	US-08-214-222-4	Sequence 4, Appl
40	123.5	7.8	289	2	US-08-467-852A-5	Sequence 5, Appl
41	123.5	7.8	289	2	US-08-468-718-4	Sequence 4, Appl
42	123.5	7.8	289	2	US-08-247-491A-5	Sequence 5, Appl
43	123	7.7	288	3	US-08-312-849-4	Sequence 4, Appl
44	123	7.7	288	3	US-08-446-201-4	Sequence 4, Appl
45	123	7.7	288	4	US-09-308-375-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-961-083-34
Sequence 34, Application US/08961083
Patent No. 6159469

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Aptamers and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-34

Query Match 57.7%; Score 916; DB 3; Length 213;
Best Local Similarity 93.6%; Pred. No. 1e-67;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKSGEGADLSMKGDVTEHQFTEQVKSNSAQOVLNMTIOVFEKQYSELDKREVD 90
|||||
1 SKSGEGADLSMKGDVTEHQFTEQVKSNSAQOVLNMTIOVFEKQYSELDKREVD 60

OY 91 TIAEEKOYGENYORVLSOAGMTLETETRAKQIRTSKLVLEAVKVAEELTDEAYKKAFDE 150
 DB 61 TIAEEKOYGENYORVLSOAGMTLETETRAKQIRTSKLVLEAVKVAEELTDEAYKKAFDE 120
 OY 151 YTPDVTQAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210
 DB 121 YTPDVTQAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTE 180
 OY 211 VP-EDYKKAFA-----LDVD 225
 DB 181 VPGASPKPLFAFRGCMVFLD 203

RESULT 2
 US-09-536-784-34
 ; Sequence 34, Application US/09536784
 ; Patent No. 6534082
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/536,784
 ; FILING DATE: 30-Oct-1997
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Michelle S. Marks
 ; REGISTRATION NUMBER: 41,971
 ; REFERENCE/DOCKET NUMBER: PB340P3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 213 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
 US-09-536-784-34

Query Match 57.7%; Score 916; DB 4; Length 213;
 Best Local Similarity 93.6%; Pred. No. 1e-67;
 Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

OY 31 SKSEGDALISMKGDTTEHOFTTEQVKSNSAOQVLLNMTIOKVFEEQYSELDDEKVD 90
 DB 1 SKSEGDALISMKGDTTEHOFTTEQVKSNSAOQVLLNMTIOKVFEEQYSELDDEKVD 60
 OY 91 TIAEEKOYGENYORVLSOAGMTLETETRAKQIRTSKLVLEAVKVAEELTDEAYKKAFDE 150
 DB 61 TIAEEKOYGENYORVLSOAGMTLETETRAKQIRTSKLVLEAVKVAEELTDEAYKKAFDE 120
 OY 151 YTPDVTQAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210
 DB 121 YTPDVTQAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDSASTE 180

OY 211 VP-EDYKKAFA-----LDVD 225
 DB 181 VPGASPKPLFAFRGCMVFLD 203

RESULT 3
 US-09-107-532A-6355
 ; Sequence 6355, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-8277
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 6355:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 344 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (B) LOCATION 1...344
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6355:
 US-09-107-532A-6355

Query Match 39.9%; Score 634; DB 4; Length 344;
 Best Local Similarity 40.5%; Pred. No. 2.6e-44;
 Matches 130; Conservative 75; Mismatches 106; Indels 10; Gaps 5;

OY 5 LKGVDM--KKKLAGATLLSVATLAACSGSEADLISMKGDTTEHOFTTEQVKSNS 61
 DB 2 ITGQILMKKKSITLATSALAVITLAACS-GDTNKDIATMKGTTIVSFDYKALESS 60
 OY 62 AQQVLLNMTIOKVFEEQYSELDDEKVDYTPDVTQAQIIRLNNEDEKAEVLEKAKAEGAD 121
 DB 61 NQSLVQRMITTYKVFNNKYGDKVTDQYDAVDQASLSGTFPSQLEAAGYTRDYKEYL 120
 OY 122 RTSKLVLEAVKVAEELTDEAYKKAFDEYTPDVTQAQIIRLNNEDEKAEVLEKAKAEGAD 181
 DB 121 RNILAFEGAIK--AHVDITDDDLTKAWKSPHVEAQIIRLSSSEDEKADV-KRSADGD 177

OY	182	F A O L A K N S D E R K T E K G S I T P D S E Y B Q V K K A A F L D V D G S V I T A G C Q A Y S S	241
		: : :	
D b	178	F S K I A K S T D T E K E K D G K V N K F D S T T T I P A V K V K E A F K L K G E L S D V I T T T T N P S V A T	237
OY	242	Q Y V I V K L T K K T E K S N I D D Y K E K L K T V I L K O K N D S T P V O S I G K E J O A N I K V K Q A R O	301
		: :	
D b	238	E Y I V K V K W K K O N K G N D M K T K D Q K D I A T E T K L S D N A F T T K V I G E L K D A N V A K I K D A E	297
OY	302	N I F P O Y I G G D S S S S S S T S N E	322
		: : : : : : : :	
D b	298	N V L S A F T ---T T S S S T K D S E	315

```

1      RESULT 4
2      US-09-071-035-182
3      : Sequence 182, Application US/09071035
4      : Patent No. 6448043
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Gil H. Choi
8      : TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
9      : NUMBER OF SEQUENCES: 496
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: Human Genome Sciences, Inc.
12     : STREET: 9410 Key West Avenue
13     : CITY: Rockville
14     : STATE: Maryland
15     : COUNTRY: USA
16     : ZIP: 20850
17     :
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage
20     : COMPUTER: HP Vectra 486/33
21     : OPERATING SYSTEM: MSDOS version 6.2
22     : SOFTWARE: Ascii Text
23     :
24     : CURRENT APPLICATION DATA:
25     : APPLICATION NUMBER: US/09/071.035
26     : FILING DATE:
27     : CLASSIFICATION:
28     : PRIOR APPLICATION DATA:
29     : APPLICATION NUMBER:
30     : FILING DATE:
31     :
32     : ATTORNEY/AGENT INFORMATION:
33     : NAME: A. Anders Brookes
34     : REGISTRATION NUMBER: 36,373
35     : REFERENCE/DOCKET NUMBER: PB369P2
36     : TELECOMMUNICATION INFORMATION:
37     : TELEPHONE: (301) 309-8504
38     : TELEFAX: (301) 309-8512
39     : INFORMATION FOR SEQ ID NO: 182:
40     : SEQUENCE CHARACTERISTICS:
41     : LENGTH: 342 amino acids
42     : TYPE: amino acid
43     : STRANDEDNESS: single
44     : TOPOLOGY: linear
45     :
46     : MOLECULE TYPE: protein
47     :
48     : US-09-071-035-182

```

Query Match	38.1%	Score	605.5	DB 4	Length	342
Best Local Similarity	41.8%	Pred. No.	5,7e-42			
Matches	132	Conservative	62	Mismatches	113	Indels
					9	Gaps
					5	
QY	10	MKKKLLIAGATILLVSATLAAAGSGSGEGADLSMKGDVITEHQFYDYQVKSNPSSAQOVLNLM	69			
Db	1	MKKKLLIILAAAGAMAFVSLAACSSGSK--DIATMKSTIVVDFFYNQIKKEOSTSQAQFSQM	58			
QY	70	TIQKFEKQYSGELDDKEVDDTIAPEKKQ---YGENYQRYVLSQAQMTLETAKQAQIRTSLL	126			
Db	59	VYIKFKEEYGGKRVDDKXIQKNFEDAKKEQVEAGSKFSDALQAGLTETETFEKKOLKORRA	118			
QY	127	VELAVKAAAEALDTAEVAKKAFDEKTPDVTAAQIILINNEDKAEVLEKAKAGADPAOLA	186			
Db	119	YDAGLK-AHLKLTDEDLKTAWASHPEVEAQIIQVASDDAKAV-KKRIITDGGPFTKA	175			
QY	187	KDNSTDEKTKENGGEITPDSASTVEPYQVKKAAFLDNDGVSVDVITATCTQAYSSQYITV	246			

Db 176 KESTSTATKAKKOGGKIKKEDSDQATYTPAEKKEAFLKNDGEVSEPIIAFTNNQTYTITYVV 235

QY 247 KLIKTEKSSNIDYIEKLTIVILQKONDSTFVSIIGKELQAAINTKVKDQAFONITFO 306

Db 236 KMKRNKAKGDMKPYEKEIKKIAEFTKLDQFVSKVSIYDELKAAVKKIKDQAFKNLAG 295

QY 307 YIGGDSSSSSSTNS 322

Db 296 YM-QTESSASSSEKKE 310

```

1      RESULT 5
2      US-09-107-532A-6113
3      : Sequence 6113, Application US/09107532A
4      : Patent No. 6583275
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Lynn A Doucette-Stamm and David Bush
8      : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
9      : ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
10     :
11     : NUMBER OF SEQUENCES: 7310
12     :
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: GENOME THERAPEUTICS CORPORATION
15     : STREET: 100 Beaver Street
16     : CITY: Waltham
17     : STATE: Massachusetts
18     : COUNTRY: USA
19     : ZIP: 02354
20     :
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: CD-ROM ISO9660
23     : COMPUTER: PC
24     : OPERATING SYSTEM: <Unknown>
25     : SOFTWARE: ASCII
26     :
27     : CURRENT APPLICATION DATA:
28     : APPLICATION NUMBER: US/09/107,532A
29     : FILING DATE: 30-Jun-1998
30     :
31     : PRIOR APPLICATION DATA:
32     : APPLICATION NUMBER: 60/085,598
33     : FILING DATE: 14 May 1998
34     : APPLICATION NUMBER: 60/051571
35     : FILING DATE: July 2, 1997
36     :
37     : ATTORNEY/AGENT INFORMATION:
38     : NAME: Arinello, Pamela Deneke
39     : REGISTRATION NUMBER: 40,489
40     : REFERENCE/DOCKET NUMBER: GTC-012
41     :
42     : TELECOMMUNICATION INFORMATION:
43     : TELEPHONE: (781)893-5007
44     : TELEFAX: (781)893-8277
45     :
46     : INFORMATION FOR SEQ ID NO: 6113:
47     : SEQUENCE CHARACTERISTICS:
48     : LENGTH: 303 amino acids
49     : TYPE: amino acid
50     : TOPOLOGY: linear
51     : MOLECULE TYPE: protein
52     : HYPOTHETICAL: YES
53     : ORIGINAL SOURCE:
54     : ORGANISM: Enterococcus faecium
55     :
56     : FEATURE:
57     : NAME/KEY: misc_feature
58     : LOCATION: (B) LOCATION 1...303
59     : SEQUENCE DESCRIPTION: SEQ ID NO: 6113:
60     :
61     : US-09-107-532A-6113

```

	Query Match	37.1%	Score 588.5;	DB 4;	Length 303;
	Best Local Similarity	40.5%	Pred. No. 1,2e-40;		
	Matches	125;	Conservative	69;	Mismatches 100; Indels 7; Gaps 5
QY		10	KKKLLAGATLLSVATLAACSKSGEGADLSMKGDVTEHQEYQVNSBPAOQVLNN	69	
		:	: :	:	:
		:	: :	:	:
DG		1	TKKRFLALAI-VLGTGLSGCTNAGE-KTAVSNGCITISEQEVMSDLKMGADSAVOOL	58	
QY		70	TQAKFEKQYSGELDDKEVDVTIAEKQVGNRYVSQAQMLETAKKAQRTSKLYEL	129	
		::: :	: :	:	:

Db 59 IYQVPEDEKYGVDVSTKEIDSDYDTKQKLGDSFDSQLKSAGYTEQTFKSDIKQSLAFQE 118
QY 130 AKKVAEALITDEAKAFDETPPVTAQIIRLNNDKAEVLEKAKAGADFAQLAKDN 189
Db 119 GKRC--HILTEDLKTAWESFEPEVAQIIQVASEDDAKDV--KKADGDDESKLAKDK 175
QY 190 STDEKTEKNGEITPDASTVEPEQYKKAFFALDVGVSDVITATGTQAYSQYIVTKLT 249
Db 176 STDTTKEKDGKGVKFPSTTTTTPAEVKEAFKAKDQVSDVITSTASTYTTTFFYYVKKV 235
QY 250 KTEKSSNIDYKEKKTIVILTKQNDSTFVOSITKEIQANIKYKDAFQNIFFQYIG 309
Db 236 KKNCKNDKDKYKKEIKELIATDTKLSDFONKVIQEVLDKANVKTIDKDFENVLSTFF-- 294
QY 310 GGDSSSSSS 318
Db 295 -SDSSTASS 302

RESULT 6

US-09-071-035-184
; Sequence 184, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: GIL H. CHO1
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 184:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-184

Query Match

Best Local Similarity 35.4%; Score 562.5; DB 4; Length 322;
Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;

QY 30 CSKSGEGADLSKMGVITTEHQYEVKSNPSAQOVLNLTQKVEKQYSGELDDKEVD 89
Db 1 CSSGSK--DIATKMGSTTFVDDDFYNOIKEOSTSQAFSQVYIKVPEEKYGDVYTDKXIQ 58
QY 90 DTIAEEKQ--GENTQRYVLSQAGMTLETRKAOIRTSKLVLEAVKVAEALITDEAYK 146
Db 59 KNDEAKEOVEAOGKGFSDALKQAGLTKETFKQLKORAAVDAGLK--AHLKTTDEDLKT 116

QY 147 AFDEYTPDYTAQIIRLNNDKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITPDS 206
Db 117 AWASFEPEVAQIIQVASEDDAKAV--KKEITDGDGTAKAKESTDTPATKDGKTKFPS 175
QY 207 ASTEPEQYKKAFFALDVGVSDVITATGTQAYSQYIVTKLTKEKSSNIDYKEKIK 266
Db 176 QATVPAEVKEAFKAKDQVSEPIATNMQYOTYYVKKMKKAKGNDKMPYEKEIK 235
QY 267 TYLTQKNDSTFVOSITKEIQANIKYKDAFQNIFFQYIGGGSSSSSSSNE 332
Db 236 KIAETRTADQYFVSKVISDELKAAVNTKIDAFKNAAGYM--QTESSSSASSEKKE 290

RESULT 7

US-09-107-532A-4736
; Sequence 4736, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariunello, Pamela Denke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4736:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (b) LOCATION 1...302
; SEQUENCE DESCRIPTION: SEQ ID NO: 4736:
US-09-107-532A-4736

Query Match

Best Local Similarity 28.4%; Score 450.5; DB 4; Length 302;
Matches 99; Conservative 61; Mismatches 130; Indels 5; Gaps 3;

QY 10 MKKKLAGATLTSVATLACSKSGEGADLSKMGVITTEHQYEVKSNPSAQOVLNLT 69
Db 12 LKQQLIMCISACAILLFSGCSKSN--TIATFKGGLITVQDDYDNPKNVSNNOITLLKM 69

QY 70 TIOKFEKYOSELDKEDDDTIAEKKOYGENYORVLSAQGTTTETRAQITSTKVEL 129
 Db 70 IYKRAFDDVYKEISKEVYKEYNEQIKKLGPNYKEQLKAVGOTBETTYKKLEKQMLAFQY 129
 QY 130 AVKKAVEALTEDAVYKKAFFEDTTPDYTAQIIRLNNEDKAKEWLEKAKAGADFAQLAKDN 189
 Db 130 GLK - ANVKLNDKDDLTAFMKREYEPVSTQIILFSTREEDKAKKKEAN - BGENPSKLVQAY 186
 QY 190 STDEKTEKNGEITPDSASTEYVEOVYKKAFFLDDGVSVDYATATGTOAYSSOYIYVKLT 249
 Db 187 GKNNKLKLTEDKSMNDSNPBELPTEYKKAFAFLKNGEVSDDIIPVDPPTYOOSYLVKVV 246
 QY 250 KTERKSNINDYKREKLTAYTLQKONDSFVQSIIGKELOANIKYKDAQOFONIE 304
 Db 247 KKODGSKKDYKSELEKIATEAKLDDIEFMKRTIKRVKAKNDVNTIKDPYKNIIE 301

RESULT 8
US-08-858-207A-508
; Sequence 508, Application US/08858207A
; Patent No. 6548328
; GENERAL INFORMATION:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

```

1 SOFTWARE: FastSEO for Windows Version 2.0
2
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/858,207A
5 FILING DATE: 03-MAY-1997
6 CLASSIFICATION: 435
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: 60/017670
9 FILING DATE: 14-MAY-1996
10 ATTORNEY/AGENT INFORMATION:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

?      SEQUENCE CHARACTERISTICS:          .
?      LENGTH: 74 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: NO. 6348328e
?      US-08-838-207A-508
Query Match      20.3%;  Score 322;  DB

```

```

Oy      10 MKKKLLAGATTTLSVATLAAKSGSEGADLTSMKGG
          |||||||
Db      1  MKKKLLAGATTTLSVATLAAKSGSEGADLTSMKGG

```

```

OY      70 TIOKVEEK 77
        |||||||
Db      61 TIOKVEEK 68

RESULT 9
US-09-134-001C-3811
; Sequence 3811, Application US/09134001C
; Patent No. 6580370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

```

```

; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3811

```

```

; LENGTH: 330
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3811

Query Match      13.28;   Score 210;   DB
Best Local Similarity 27.58;   Pred. No. 1.46-
Matches 95; Conservative 55; Mismatches

```

```

0Y      23 5VAT1AAC-SKSGEAD--1NSK-GVV---1IE
      |  |  |  |  |  |  |  |  |  |  |  |  |
Db      19 SALLGACGSNATESKNDTLTSSKAGDVKKADVKK
      |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      75 FEKQYSELDDKEVDDTLIAEKKQY--ENYQRYL
      :  :  :  :  :  :  :  :  :  :  :  :  :
Db      73 LADYKDKVDYDKTIDKIDKKEEKQYGGKQDFESML
      |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      130 AKKVAFAELTDAAVKKAFDFEYDPDYAOT---TR

```

```

QY 235 GTOAISQIYIVLTKTE--KSSNI-----
Db 183 EKNPNKFGEIAKESMDSSAKKOGSLGY-VIRGQ
QY 179 ---GADFAOIAKDNSTDEKTKENGELTFPDSASTE
Db 133 DKVNNSDKELKENSKK-----TSHIIIAKVV
QY 120
Db 120

```

DB 242 G-----YHIKADKETDENSESKNOKUKEER
QY 275 ND--STFVOSIIEKELOANIKVKDQAFONIFTOY
DB 295 RRIKKAIEDSILPDP-----KIKQOQOQO-----QS

US-09-252-991A-17422
Sequence 17422, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NOCLETIC ACID AND AMINO AC
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTIC
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A

```

:
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
:
: SEQ ID NO 17422

```

STUDIES RELATING TO STAPHYLOCOCCUS THERAPEUTICS

Length 330;
Tandels 84; Gaps 20;

```
-----KAKEVELEKAAE- 178  
XKSNSPSAQVLLNMTCQV 74  
      : : |  
      : : |  
      : : ||  
IANTS-----FSIVLKV 72  
      : : |  
      : : |  
      : : ||  
ETRKAQIRTS---KLVEL 129  
      : : |  
      : : |  
      : : ||  
LDDYKEOKKLSAYQKOLL 132
```

-----DDYKFKLTFTILTOHQ 274

KALFLKLEGEVSKVVKTDY 241

KAFAFDVDGVSDVI -TAT 234

EGLSDKKRAKEAKEIKQREV 182

```

      1 1 1 1 1
      :
      1
LPLTDAYKELLKEYKVDYKD 294
SSSSS 318
:::|
INSGS 330

```

REFERENCES RELATING TO PSEUDOMONAS THERAPEUTICS

[illegible][illegible]


```

Dh      39  -QSKAEKDYDAAKKQAKNNKRAVEDAQKALDANKAAQKK -YDEDOKKTJEEKAAL- ---EK 92

Qy      119 AQIRTSKIVELAKKKAABELTDEAKKAFDEITYTDVTAQIIRLNNEDKAEVLEKAAE 178
      93 A--ASEEEDKAAVAAYQAYL---AYQATDCKAARDADAKMI- ---DEAKRREEAKTK 141

Dh      179 -----GADFQALAKONSTDTEKTRKENGSEIT--FDSASTVEPQVYKKAFAALDVGVS 228

Qy      142 FNTVRAMVYVPEPQOLAEKTKKSEBAKQKRAPELTKLKEBAKKALEAKKALEAKQYDAE 201

Dh      142 FNTVRAMVYVPEPQOLAEKTKKSEBAKQKRAPELTKLKEBAKKALEAKKALEAKQYDAE 201

Qy      179 -----GADFQALAKONSTDTEKTRKENGSEIT--FDSASTVEPQVYKKAFAALDVGVS 228

Dh      142 FNTVRAMVYVPEPQOLAEKTKKSEBAKQKRAPELTKLKEBAKKALEAKKALEAKQYDAE 201

Qy      229 DVITAFGTQVAYSQYIVKLTKEKSSNIDY--KEKLTVI--LQKQNDSTFFVOSII 284
      202 EVAPQAKIAELAEQVHRLQELKEIDSESEDYAKEGFRAPLOSLDKAKKALSTIEELS 261

Dh      285 GK---ELQAAINIKYKQD 298
      262 DKIDELDAETAKLEDO 277

```

```

01 RESULT 13
02 US-08-214-164-2
03 ; Sequence 2, Application US/08214164
04 ; Patent No. 5728367
05 ; GENERAL INFORMATION:
06 APPLICANT: BRILES, DAVID E.
07 APPLICANT: YOTHER, JANET L.
08 TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
09 NUMBER OF SEQUENCES: 3
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Shoemaker and Mattare, Ltd
12 STREET: Suite 1203, 2001 Jefferson Davis Highway
13 CITY: Arlington
14 STATE: Virginia
15 COUNTRY: U.S.A.
16 ZIP: 22202-0286
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patentin Release #1.0, Version #1.25
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/214,164
24 FILING DATE: 17-MAR-1994
25 CLASSIFICATION: 424
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 07/656,773
28 FILING DATE: 15-FEB-1991
29 CLASSIFICATION: 424
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Berkstresser, Jerry W.
32 REGISTRATION NUMBER: 22,651
33 REFERENCE/DOCKET NUMBER: 6102-137
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (703) 415-0810
36 TELEFAX: (703) 521-0813
37 TELEX: LUKPAT WASHINGTON
38 INFORMATION FOR SEQ ID NO: 2:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 619 amino acids
41 TYPE: amino acid
42 TOPOLOGY: linear
43 MOLECULE TYPE: protein
44 US-08-214-164-2
45
46 Query Match 8.5%; Score 135.5; DB 1; Length 619;
47 Best Local Similarity 24.4%; Pied. No. 0.0043;
48 Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15
49
50 11 KKLLAGATITLTSVATIACSGSGGADILSKKGVITHQRYEYVKSNPSAQVILLNMT 70
51 |||:::~||| | : : : : : |
52 Db 3 KKKM--LTSLASVAI-----GAGFVASOPTVV-----RAESFPVS----- 38
53 71 IOKVEKGYSGELDD-----KEVDTIAEERKQYGGNTORVLSOAGMTLETRK 118

```

Db	39	-G\$AKEDYDAKAKKODKNNKKAVEDAKALDDPAKKAQOK--YDEDOQKTEEEKAAL---EK	92
Oy	119	AQIRTSKIVELAKKKAEBELTDEYAKKAFPEYTDVYAOIIRLNNDKAKEVYLERAAE	178
Db	93	A--ASBEMDKAAVAAYQAYL--AYQATDPAKADPAADKMI---DEAKKREFEAKTK	141
Oy	179	-----GADFAQLAKDNSTDEKTEKENGELT--FDSASTPEVBOYKKAAPALDDVGS	228
Db	142	FNTVRAMVVEPEEQLALETKKKSEBEAKOKAPBELTTLKLEBAKLEBAEKKAIEAKOKDAE	201
Oy	229	DVITATGTOYASSOYIVKLTTRKTEKSSNIDY--KEKLEKY---LTQONDSFFYOII	284
Db	202	EVAPQAKIALENOYHRLQELKEIDSESDYAKAGFRAPLOSKLDKAKKLSKLEBLS	261
Oy	285	GR--ELQANIKKQD	298
Db	262	DKIDEDPAELAKLEDO	277

```

14 RESULT 14
US-08-467-852A-3
Sequence 3, Application US/08467852A
Patent No. 5856170
GENERAL INFORMATION:
APPLICANT: BRILES, David E.
APPLICANT: YOTER, Janet L.
APPLICANT: MCDANIEL, Larry S.
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467, 852A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOMALSKI, Thomas J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454312-2064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 619 amino acids
TYPE: amino acid
STRANDEDNESS: n/a
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-467-852A-3

Query Match 8.5%; Score 135.5; DB 2; Length 619;
Best Local Similarity 24.4%; Pred. No. 0.0043;
Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15

11 KKLIIAGITLISVATTLAAGSGSGADLISKGVITETHOFYEYKSNPSAQOVLNMT 70
||||| : ||||| || : : : |||
3 KKKMI-LTSLASVAIL-----GAGFVASQPIVV-----FAESPVAS----- 38

71 IQVFEKRYGSELD-----KEYVDITAEKKYQYGENQVRLSQAQMTLETRK 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39 -QSAEKRYDARKKAKRAVEAQAALDADAKAAQKK-VDEDQKTEEKAAL-----EK 92

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:36:19 ; Search time 61 Seconds
(without alignments)
722.029 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SWTYLKGVDMKKRLAGAIT.....IFQYIGGDSSTSSSTSN 322

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 510680 seqs, 136781880 residues

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541	97.0	313	11	US-09-769-787-163
2	916	57.7	213	9	US-09-765-272-34
3	212.5	13.4	320	8	US-08-781-986A-5192
4	138	8.7	670	11	US-09-298-523B-63
5	135.5	8.5	619	11	US-09-882-774-1
6	131.5	8.3	2478	9	US-09-298-523B-3
7	130.5	8.2	2478	9	US-09-815-242-5816
8	130.5	8.2	2478	9	US-09-815-242-12967
9	130	8.2	690	11	US-09-298-523B-61
10	130	8.2	701	11	US-09-298-523B-62
11	129.5	8.2	589	11	US-09-298-523B-14
12	129.5	8.2	691	11	US-09-298-523B-1
13	129.5	8.2	707	11	US-09-298-523B-2
14	129.5	8.2	929	11	US-09-298-523B-60
15	129.5	8.2	1576	15	US-10-037-182-16

16	129.5	8.2	1609	10	US-09-938-275-11	Sequence 11, Appl
17	129.5	8.2	1609	15	US-10-037-182-14	Sequence 14, Appl
18	129.5	8.2	1609	15	US-10-299-058-12	Sequence 12, Appl
19	123	7.7	2285	10	US-09-932-183A-2	Sequence 2, Appl1
20	122.5	7.7	428	11	US-10-254-995-7	Sequence 7, Appl1
21	122.5	7.7	439	11	US-09-056-019-37	Sequence 37, Appl1
22	122.5	7.7	1786	10	US-09-742-096-3	Sequence 3, Appl1
23	119.5	7.5	530	9	US-09-159-469-73	Sequence 73, Appl1
24	119.5	7.5	530	9	US-09-159-469-73	Sequence 73, Appl1
25	119.5	7.5	550	9	US-09-159-469-64	Sequence 64, Appl1
26	119.5	7.5	550	9	US-09-159-469-64	Sequence 64, Appl1
27	117	7.4	767	9	US-09-815-242-5899	Sequence 5899, Ap
28	117	7.4	767	9	US-09-815-242-13140	Sequence 13140, A
29	114.5	7.2	439	11	US-09-056-019-28	Sequence 28, Appl
30	114.5	7.2	2368	9	US-09-815-242-5635	Sequence 5635, Ap
31	114.5	7.2	2368	9	US-09-815-242-12389	Sequence 12389, A
32	114	7.2	431	12	US-10-254-995-3	Sequence 3, Appl1
33	114	7.2	481	11	US-09-298-523B-6	Sequence 6, Appl1
34	113.5	7.1	1167	9	US-09-815-242-11522	Sequence 11522, A
35	112.5	7.1	425	12	US-10-254-995-13	Sequence 13, Appl
36	112.5	7.1	437	11	US-09-056-019-32	Sequence 32, Appl
37	112	7.1	841	9	US-09-861-451A-30	Sequence 30, Appl
38	111.5	7.0	424	12	US-10-254-995-14	Sequence 14, Appl
39	111.5	7.0	437	11	US-09-056-019-29	Sequence 29, Appl
40	111.5	7.0	487	11	US-09-298-523B-9	Sequence 9, Appl1
41	111	7.0	428	12	US-10-284-400-8	Sequence 8, Appl1
42	111	7.0	429	12	US-10-284-400-16	Sequence 16, Appl
43	111	7.0	466	11	US-09-298-523B-4	Sequence 4, Appl1
44	111	7.0	764	15	US-10-309-851-10	Sequence 10, Appl
45	111	7.0	1133	15	US-10-309-851-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-769-787-163
; Sequence 163, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Phillip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 163
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-163

Query Match 97.0%; Score 1541; DB 11; Length 313;
Best Local Similarity 100.0%; Pred. No. 2.3e-107;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 MKKLLGATLTLVATLAACSKSGEGADLSKMGDVTTERHOFEDYKSPSAQOVLNM 69
DB 1 MKKLLGATLTLVATLAACSKSGEGADLSKMGDVTTERHOFEDYKSPSAQOVLNM 60
QY 70 TIOQVEKQYSGSELDKEDVDITAEKKQYGENORVLSQAGMTLERRKOITRSKVEL 129
DB 61 TIOQVEKQYSGSELDKEDVDITAEKKQYGENORVLSQAGMTLERRKOITRSKVEL 120
QY 130 AVKVAEAEITDEAYKKAFDEYTPDVYQAIIRLNEDKAEVLEKAKAEAGADPAQLAKDN 189

Db 121 AVKVAEAEITDEAYKKAFFDEYTPDVTQAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDN 180
QY 190 STDEKTEKNGEITFDSASTEVPEQVKAAPALPDVGVSDVITATGTAQVSSOYIYKLT 249
Db 181 STDEKTEKNGEITFDSASTEVPEQVKAAPALPDVGVSDVITATGTAQVSSOYIYKLT 240
QY 250 KTEKSSNIDYKKEKLTVITITOKONDSTFVOSIIGELQANIKVKDQAFQNIETQYIG 309
Db 241 KTEKSSNIDYKKEKLTVITITOKONDSTFVOSIIGELQANIKVKDQAFQNIETQYIG 300
QY 310 GGDSSSSSTSTNE 322
Db 301 GGDSSSSSTSTNE 313

RESULT 2
US-09-765-272-34
Sequence 34, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-765-272-34

Query Match 57.7%; Score 916; DB 9; Length 213;
Best Local Similarity 93.6%; Pred. No. 5,9e-61;
Matches 150; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

QY 31 SKSSEADLSMGDVITTEHOFEYOYKSNPSAOVLNMTIQVFEKQSGSELDDEVD 90
Db 1 SKSSEADLSMGDVITTEHOFEYOYKSNPSAOVLNMTIQVFEKQSGSELDDEVD 60
QY 91 TIAEKKQYGENQVYLSQAGMTLETTRKQIIRTSKVELAVKVAEAEITDEAYKKAFFDE 150
Db 61 TIAEKKQYGENQVYLSQAGMTLETTRKQIIRTSKVELAVKVAEAEITDEAYKKAFFDE 120
QY 151 YTPDVTQAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210
Db 151 YTPDVTQAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210

Db 121 YTPDVTQAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 180
QY 211 VP-EQVKAAPAF-----LDVD 225
Db 181 VPGASPKRPLFAFRGMYFLDVD 203

RESULT 3
US-08-781-986A-5192
Sequence 5192, Application US/08781986A
Publication No. US2003005436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 5192:

SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-781-986A-5192

Query Match 13.4%; Score 212.5; DB 8; Length 320;
Best Local Similarity 23.5%; Pred. No. 3.2e-08;
Matches 72; Conservative 70; Mismatches 107; Indels 57; Gaps 12;

QY 23 SVATLACSGSGGAD-----LSMGDVITTEHOFEYOYKSNPSAOVLNMTIQVFEK 77
Db 14 SALLTACG-GASATSKENTLISKAGDVTADTKKIKGDKDIANASPTKM-LNKILAD 70
QY 78 QYSELDKKEVDPTIAEKKQYGENQVYLSQAGMTLETTRKQIIRTSKVELAVKVAEAEITDEAYKKAFFDE 132
Db 71 KYNNKNDKKIDQIEKMKQYGGKDFERAKLQOQGLTADKYKENTRTAHHKELLSDKI 130
QY 133 KVAEAEITDEA-----YKKAFFDEYTPDVTQAQIIRLNNEKAKAEVLEKAKAGAD 181
Db 131 KIDSSEIKEDSKASHILIKVSKSKDXEGLDDKEA-----KKAFFDEYTPDVTQAQIIRLNNEKAKAEVLEKAKAGAD 163
QY 182 FAQLAKDNSTDEKTEKNGEITFDSASTEVPEQVKAAPALPDVGVSDVITATGTAQVSS 241
Db 184 FGEIARKKESMDGSAKDELQY-VLKGQTDKDFERAKLKGVESEVYKS-----SF 236
QY 242 QYIYVLTFRKTEKSSNIDYKKEKLTVITITOKONDSTFVOSIIGELQANIKVKDQAFQNIETQYIG 301
Db 237 GYHIIKADKPTDPSNSKOSLKEKL-----VDQVQV-KNKKLLTDAK 277

QY 302 NITFOY 307
DB 278 DLKEY 283

RESULT 4
US-09-298-523B-63

; Sequence 63, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-63

Query Match 8.7%; Score 138; DB 11; Length 670;

Best Local Similarity 20.6%; Pred. No. 0.032; Matches 85; Conservative 56; Mismatches 138; Indels 134; Gaps 14;

QY 31 SKGSEADLISMKGDVITEHOFYEQVKNPSAQOVLNMTOKVFEKQY-----80
DB 181 SKKAEATRLLEIKTE---RKKAEEBAKRAESEEKKAEEKAKOVDTKEGCKPKRRAKGV 237
QY 81 -----SELDDEKVDITIAE-----KKOYGENYQVLSQ-----109
DB 238 SGELATPDKKENAKSSDSSVGEETLPSPSLMANANESOTHRKDVEDYIKMLSEIQDLR 297
QY 110 ----AGMTLETRKAQIRTSKLVELAV-----K 132
DB 298 RKHTQNNVNLIKLSAIRTLYELSVLEKNSKEELTSKKAELTAFAEQFKKDTLKEPK 357
QY 133 KVAEAE-LTDEAYKKAFDEYTPD-----VTAQIIRL 162
DB 358 KVAEAEKKEVEEAKKADOKEDRNRNPTNYTKTLELAESDVAKVDAELELYVEEANES 417
QY 163 NNEDEKAEVLEKAKAEADFAQAKONSTDEKTEKNGEITFPDSASTEVPEQYKAFAL 222
DB 418 RNEEKIQAKKEKESKAEATRLKIKITDOKKAE-----EEAKKRAESEEKKAEEK 470
QY 223 D-VDGVADVTATGTOAYSSQYIVLTKKTEK--SSNIDY-KEKLTQVI---LTOKON 275
DB 471 QKVDAAEYALAE---KIAELEFYQRLKEKELEDESDSEYKELRAPLOSKLDTRKA 527
QY 276 DSTFVQSIIGK-----ELQANIKVKDQAFONITFOYIGGDSSSSSSTSN 322
DB 528 KLSKLEELSKRIDELDAEIAKLEVQLKDAEGNNVVEAYFKEGKTTAEKKA 580

RESULT 5
US-09-882-774-1

; Sequence 1, Application US/09882774
; Publication No. US20030021795A1
; GENERAL INFORMATION:
; APPLICANT: Houston, Michael E.
; APPLICANT: Hodges, Robert
; TITLE OF INVENTION: Use of Coiled-Coil Structural Scaffold to Generate
; FILE REFERENCE: 003592-007
; CURRENT APPLICATION NUMBER: US/09/882,774
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,892
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/213,387

; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 1
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-882-774-1

Query Match 8.5%; Score 135.5; DB 11; Length 619;

Best Local Similarity 24.4%; Pred. No. 0.044; Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15;

QY 11 KKKLLGATILLVATLAACSGSEADLISMKGDVITEHOFYEQVKNPSAQOVLNMT 70
DB 3 KKKMI--LTLASAVAIL-----GAGFVASQPTV-----RAEESVSA-----38
QY 71 IQVFEKQYSELD-----KEYDVTIAEKKOYGENYQVLSQAGMTLETRK 118
DB 39 -QSKAEKDYDAKKAKDAKNAKAVEDAQKALDDAKAAQK-YDEDDOKTEEKAAL---EK 92
QY 119 AQIRTSKLVELAKKVAEALITDEAYKKAFDETPVTOIIRLNNEDEKAEVLEKAKAE 178
DB 93 A---ASEEMDKAVAAVQOAYL---AYQOATDKAKADAADM---DEAKKREBEAKTK 141
QY 179 -----GADFAOLAKDNSTDEKTEKNGEIT--FDSASTEVPEQYKAFALVDVGS 228
DB 142 FNTVIRAVNYPEPEQIAETKRSSEAKOKAPELTKLEAKAKIEAKKATEKOKVDAE 201
QY 229 DVITATGTOAYSSQYIVLTKKTEKSSNIDY-KEKLTQVI---LTOKONSTFVOSIT 284
DB 202 EVAPQAKIAELENOVHRLFEQLEKIDESSEYAKEGFAPLOSKLDKAKKLSKLEELS 261
QY 285 GK--ELQANIKYKQD 298
DB 262 DKIDELDAEIAKLEDDQ 277

RESULT 6

US-09-298-523B-3

; Sequence 3, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-3

Query Match 8.3%; Score 131.5; DB 11; Length 711;

Best Local Similarity 20.7%; Pred. No. 0.11; Matches 86; Conservative 55; Mismatches 137; Indels 137; Gaps 15;

QY 31 SKGSEADLISMKGDVITEHOFYEQVKNPSAQOVLNMTOKVFEKQY-----80
DB 221 SKKAEATRLLEIKTE---RKKAEEBAKRAESEEKKAEEKAKOVDTKEGCKPKRRAKGV 277
QY 81 -----SELDDEKVDITIAE-----KKOYGENYQVLSQ-----109
DB 278 SGELATPDKKENAKSSDSSVGEETLPSPSLMANANESOTHRKDVEDYIKMLSEIQDL 337
QY 110 ----AGMTLETRKAQIRTSKLVELAV-----K 131
DB 338 RKHTQNNVNLIKLSAIRTLYELSVLEKNSKEELTSKKAELTAFAEQFKKDTLKEPK 397

QY 28AACGSGSGAGDIAISKMGVITEH-----QFEYOVSNSQAQVLNMTIQKVFEPKOYGSL 83
11 :: : :: 11 ::
Db 1569 AAYEGGLNNINATTGGVTAKTAVAQVKOOLHAND-----YKPAGK-- 1612
QY 84 DKEVDPTIAEKKOGEN--YORVLSQAGMTELETKRAAIRSKLVELAVKKVAEELT 140
11 :: : :: 11 ::
Db 1613 --KELDQAAARKKQIOETPNASQEIEINDAKOEVDTELNOAKTN-----VOOSITNEAV 1664
QY 141 DEAYK-----KAPEYTPDTTAOI-----IRLNEDK-----AKEYLEKAKREGAD 181
11 :: : :: 11 ::
Db 1665 DNAVEGKATINAKYTSEYKKDLAKLIEDIYNNAKNKWEADNSASTSSETAE-AKOKLAE 1722

QY	28	AACSGSEGADLI	MKGGVITEH----	-QFYDQVSNPAAQVLNMTIÖKPEKROYSEL	83	
Db	1569	AAYEEGLNNINAAITTTGGDYVTAKTAPAVOQÜOLANP-----	-----VKIPAGK--	16117		
QY	84	DKEEVDDTIAEEKKOYSEN---	YORVISQACMTELETKRAQIRTSKYELAVAKVAEAELT	140		
Db	1613	--KELDDAAMADKKTOIEDTPWASOOEIENDAKOEVDTELNOÄKTN-----	-VDOSSTNEYV	166		
QY	141	DEAYK-----	KAFDEYTPDVTAQI-----	IRLNNEEK-----	AKEYELEKAKAEAD	181
Db	1665	DNAVEEKAKINAVKITISEYKKDLAIETDAYNARKVANADSNSASTSEIAE-AKÖKLAÆ	17233			
QY	182	FAQLANDNSTDEKTEKENGELETTFEDSASTVEQEYKKAALFADVDGSVPVITATG-TQAYS	240			
Db	1724	LKQTADGNVQAANSKD-----	DIEVOÜHN-----	DDINDINTYPIPTCKESAT	1766	
QY	241	SÖYIYVYLKTKTESSNIDDIKEKLKVILLOKONDSSTPOVOST	283			
Db	1767	TDLIYAAYADOKKNINSADTNATOUEKQOAIKOVDPNOVTALESII	1809			

Db 454 AESSEKAAAEKQKVADEEYALEA--KIALEVEVQRLKEKEIDESDEYLKRGRL 510
 QY 267 TVI---LTKQNDSTFVQSIIGK-----ELQANIKVKDAQFQNIPTQYIGGSSSS 317
 Db 511 APLQSKLDTKKAKLSKEELSDKIDELDAELAKLEVLQAKAEGNNVEALFKRGLEKTTA 570
 QY 318 STSNE 322
 Db 571 EKKA 575

RESULT 12

US-09-298-523B-1

; Sequence 1, Application US/09298523B
 ; Publication No. US20030059438A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/298,523B
 ; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 691
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-298-523B-1

Query Match

8.2%; Score 129.5; DB 11; Length 691;
 Best Local Similarity 22.5%; Pred. No. 0.14;
 Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVFQKQYSGELDKEDVDITAE 95
 Db 269 VSGELATPPKKENDAKSSDSVGEETLPSPSLMANESQTEHR-----KDYDEYI--- 318
 QY 96 KKQYGE-NYQRYLSQAGMTLETRKAOIRSKVLELAV----- 131
 Db 319 KKMISEIQLDRKHTQNVNINIKLSAIFKYYLELVLEKNSKKEELTSKTAELTAFAE 378
 QY 132 -----KKVAEAE-LTDEAYKKADEYTPD----- 154
 Db 379 QKFKDLPKPEKKVAEKKVEEKKAKQDEEDRRNYPNTYKTLLEIAESDVYKKA 438
 QY 155 ----VTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNDSTDEKTEKNGEITFDSASTE 210
 Db 439 EELVLEKANESENEKIKQAKKEVSKKAFAETRLER-ITDKKAAE-----EAKKK 490
 QY 211 VPEQVKAAPALD-VGVSDVITATGTQAYSSQYIVKLTQTEK--SSNIDY-KKKL 266
 Db 491 AESEKKAFAKQKQVDAEEYALEA--KIALEVEVQRLKEKEIDESDEYLEKGLR 547
 QY 267 TVI---LTKQNDSTFVQSIIGK-----ELQANIKVKDAQFQNIPTQYIGGSSSS 317
 Db 548 APLQSKLDTKKAKLSKEELSDKIDELDAELAKLEVLQAKAEGNNVEALFKRGLEKTTA 607
 QY 318 STSNE 322
 Db 608 EKKA 612

RESULT 13

US-09-298-523B-2

; Sequence 2, Application US/09298523B
 ; Publication No. US20030059438A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/298,523B

; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 707
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-298-523B-2

Query Match
 8.2%; Score 129.5; DB 11; Length 707;
 Best Local Similarity 20.9%; Pred. No. 0.15;
 Matches 77; Conservative 57; Mismatches 122; Indels 113; Gaps 14;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVFQKQYSGELDKEDVDITAE 99
 Db 269 VSGELATPPKKENDAKSSDSVGEETLPSPSLMANESQ-----TEHRKDY 314
 QY 100 GENYQVLSQ-----AGMTLETRKAOIRSKVLELAV----- 131
 Db 315 DEYIKKMLSEIQLDGRKHTPNTNINIKLSAIFKYYLELVLEKNSKKEELTSKTAELT 374
 QY 132 -----KKVAEAE-LTDEAYKKADEYTPD----- 154
 Db 375 AAFQFKDLPKPEKKVAEKKVEEKKAKQDEEDRRNYPNTYKTLLEIAESDVYK 434
 QY 155 ----VTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNDSTDEKTEKNGEITFDS 206
 Db 435 VKAELELVKEANESENEKIKQAKKEVSKKAFAETRLER-ITDKKAAE-----E 486
 QY 207 ASTEVEQVKAAPALD-VGVSDVITATGTQAYSSQYIVKLTQTEK--SSNIDY-K 262
 Db 487 AKKAESEKKAFAKQKQVDAEEYALEA--KIALEVEVQRLKEKEIDESDEYLYK 543
 QY 263 EKLKTYI---LTKQNDSTFVQSIIGK-----ELQANIKVKDAQFQNIPTQYIGGDS 313
 Db 544 EGRAPLQSKLDTKKAKLSKEELSDKIDELDAELAKLEVLQAKAEGNNVEALFKRGLE 603
 QY 314 SSSSTSTNE 322
 Db 604 KTTAEKKA 612

RESULT 14

US-09-298-523B-60

; Sequence 60, Application US/09298523B
 ; Publication No. US20030059438A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/298,523B
 ; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 60
 ; LENGTH: 929
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-298-523B-60

Query Match
 8.2%; Score 129.5; DB 11; Length 929;
 Best Local Similarity 22.5%; Pred. No. 0.21;
 Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVFQKQYSGELDKEDVDITAE 95
 Db 269 VSGELATPPKKENDAKSSDSVGEETLPSPSLMANESQTEHR-----KDYDEYI--- 318
 QY 96 KKQYGE-NYQRYLSQAGMTLETRKAOIRSKVLELAV----- 131
 Db 319 KKMISEIQLDRKHTQNVNINIKLSAIFKYYLELVLEKNSKKEELTSKTAELTAFAE 378

Search completed: August 29, 2003, 10:45:39
Job time : 62 secs

[illegible]

```

RESULT 15
US-10-037-182-16
Sequence 16, Application US/10037182
Publication No. US20030044899A1
GENERAL INFORMATION:
APPLICANT: Tivvasson, Karl
APPLICANT: Doi, Masayuki
APPLICANT: Thyboll, Jill
TITLE OF INVENTION: Recombinant LamnIn 10
FILE REFERENCE: 99-274-F
CURRENT APPLICATION NUMBER: US/10/037,182
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,449
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/279,282
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 1576
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-182-16

```

Query Match	8.2%	Score 129.5; 4	DB 15;	Length 1576;
Best Local Similarity	20.9%	Pred. No. 0.44;		
Matches	65;	Conservative	136;	Indels 37; Gaps 13

```

OY 32 GSEADIDISKMG---DYTHEOFPEOVKSNPSAQOVLNMTIOKVEFKOYGSLEDC-- 86
Db 1015 KLOEELSIANLNGIDGEMVTDQAFEDRK---EAEREYMDLLREQDYKVDQNLMDLQ 1071
OY 87 EVDDTIAEEKOYGENVOVLISQAGMTLETBRAOI-RTSKVELAVKVAEALTDYAKV 145
Db 1072 RVNNTLSSQISRL-ONIRNTIETGNGLAEQARAHVENERTLEIETISRELEKAKVAANVS 1130
OY 146 KAFDEYTPDYAQOITRLNN---EDKAEVELEKAAEGADPAQOLAK---DNSTDE----- 193
Db 1131 VTQPESTDp-----NNMTLLAEAKKLBRHKQEDDLYRAKTAINDISTEYNNLL 1185
OY 194 KTKENGEEITPD-SASTEVPQOVKKAAPALDVG--VSDVTYANGTQA--YSSOYIYVK 247
Db 1184 RTLGENGNTAEIEIENLRKYEAKNISODLEQOARVHEEAKRAGDKAVEIYASVQSLP 1243
OY 248 LTKRT--EKSSNIDYREKLTIVLITOKONSTFVQSIGLELQANAKVVDQAOQNFT 305
Db 1244 LDSEIIEANNNIKHEALENLEOLIQKLKDYEDLREDMRKGELEVKNNLEKCKTQQTAD 1305
OY 306 QYIGGDSSSS 316
Db 1304 QLLARADAKA 1314

```

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:31:09 ; Search time 41 Seconds
(without alignments)
755.276 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SMNYLKGVMKKLAGAIF.....IFQYIGGDDSSSSSTSN 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	97.0	313	2	E95113
2	1538	96.9	313	2	D97982
3	512	32.2	308	2	E86840
4	400.5	25.2	291	2	AF1722
5	393.5	24.8	293	2	AC1352
6	375.5	23.6	299	2	A44858
7	373	23.5	299	2	S08083
8	367	23.1	299	2	A32313
9	357.5	22.5	292	2	S15269
10	328	20.7	294	2	AD1255
11	327.5	20.6	294	2	A11617
12	286	18.0	333	2	A83797
13	212.5	13.4	320	2	H89970
14	189.5	11.9	353	2	D97295
15	187	11.8	677	2	E75563
16	169.5	10.7	594	2	B64018
17	164.5	10.4	623	2	A64774
18	163.5	10.3	623	2	G90690
19	163.5	10.3	623	2	G85541
20	162	10.2	628	2	A10382
21	158.5	10.0	623	2	AG0558
22	158	9.9	619	2	E82141
23	156.5	9.9	297	2	S66102
24	155.5	9.8	282	2	E71662
25	155.5	9.8	336	2	D70113
26	151	9.5	282	2	H97809
27	151	9.5	331	2	AF3267
28	146	9.2	273	2	S52412
29	142.5	9.0	430	2	B83572

30	140.5	8.8	314	2	A87629	hypothetical prote
31	139.5	8.8	437	2	B70353	conserved hypothet
32	135.5	8.5	619	2	A41971	surface protein ps
33	135.5	8.5	619	2	A97887	surface protein ps
34	135	8.5	237	2	F70446	conserved protein ps
35	135	8.5	1526	2	A45605	mature-parasite-in
36	134	8.4	358	2	AB1746	hypothetical prote
37	133.5	8.4	888	2	T25713	hypothetical prote
38	132.5	8.3	313	2	F64069	hypothetical prote
39	131.5	8.3	623	2	G84985	peptidyl-prolyl iso
40	131	8.2	247	2	A96934	peptidyl-prolyl ci
41	131	8.2	624	2	PC6003	surface membrane p
42	131	8.2	979	2	B90601	hypothetical prote
43	130	8.2	329	2	G72319	hypothetical prote
44	130	8.2	2139	2	T18296	myosin heavy chain
45	129.5	8.2	1609	1	MMHDB2	laminin gamma-1 ch

ALIGNMENTS

RESULT 1
E95113
proteinase maturation protein, probable [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: E95113
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewys, M.R.; Radune, D.; Holtzappl
son, T.; Hickey, E.R.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: E95113
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <KUR>
A:Cross-references: GB:AE005672; PIDN:AK75102.1; PID:914972456; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
A:Genetics:
A:Gene: SP0981

Query Match	97.0%;	Score 1541;	DB 2;	Length 313;
Best Local Similarity	100.0%;	Pred. No. 3.3e-74;		
Matches 313;	Conservative	0;	Mismatches 0;	Indels 0;
QY	10	MKKKLGAATLLTAAATLAACSKSGEGADLISMGDVTTEHQFYEQVKNPSAQOVLNM	69	
DB	1	MKKKLGAATLLTAAATLAACSKSGEGADLISMGDVTTEHQFYEQVKNPSAQOVLNM	60	
QY	70	TICKVFEKQYSGELDDKEVDVDTIAEKKQYGENYQVLSQAGMTLEFRKQIRTSKVEL	129	
DB	61	TICKVFEKQYSGELDDKEVDVDTIAEKKQYGENYQVLSQAGMTLEFRKQIRTSKVEL	120	
QY	130	AVKVAFAELTDEAYKKAPEYTPDVTAQIIRLNNEKAEVEKAAEGADFAQLAND	189	
DB	121	AVKVAFAELTDEAYKKAPEYTPDVTAQIIRLNNEKAEVEKAAEGADFAQLAND	180	
QY	190	STDEKTKENGELTFDSASTVEPQVKKAFALDQVSDVITATGQAVSSQYIYKLT	249	
DB	181	STDEKTKENGELTFDSASTVEPQVKKAFALDQVSDVITATGQAVSSQYIYKLT	240	
QY	250	KTEKSSNIDDKKELKLVITLTKQNDSTFVQSTIGKELQANIKVQDAFQNTFYIG	309	
DB	241	KTEKSSNIDDKKELKLVITLTKQNDSTFVQSTIGKELQANIKVQDAFQNTFYIG	300	
QY	310	GGDSSSSSSTSNE 322		
DB	301	GGDSSSSSSTSNE 313		

RESULT 2

RESULT 2

D97982
Proteinase maturation protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97982
R: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A: Reference number: A97872; MUID:2143245; PMID:11544234
A: Accession: D97982
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-313 <KUR>
A: Cross-references: GB:AE007317; PIDN:AAK9688.1; PID:q15458489; GSPDB:GN00174
C: Genetics:
A: Gene: pmpa

Query Match 96.9%; Score 1538; DB 2; Length 313;
Best Local Similarity 99.7%; Pred. No. 4,7e-74;
Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY 10 MKKLLAGATLLSVATLAACSGSEGADLISMKGDIYTHQFEQVKNPSAQOVLNM 69
    |||||
Db 1 MKKLLAGATLLSVATLAACSGSEGADLISMKGDIYTHQFEQVKNPSAQOVLNM 60
QY 70 TIQKFEKQYGESELDKEVDVDTIAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVEL 129
    |||||
Db 61 TIQKFEKQYGESELDKEVDVDTIAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVEL 120
QY 130 AVKVAEALTDKAEKKAFFDEYTPDVTAOIIRLNNEKAEVLEKAKAEADPAQLAKDN 189
    |||||
Db 121 AVKVAEALTDKAEKKAFFDEYTPDVTAOIIRLNNEKAEVLEKAKAEADPAQLAKDN 180
QY 190 STDKTEKNGEITFDSASTVEPEQVKAFAFDVGVSVITATGQAVSQDYIYVLT 249
    |||||
Db 181 STDKTEKNGEITFDSASTVEPEQVKAFAFDVGVSVITATGQAVSQDYIYVLT 240
QY 250 KTEKSSNIDYKKEKTKTVILTKQNDSTFVOSITKELOANIKYKDAQFONIFQYIG 309
    |||||
Db 241 KTEKSSNIDYKKEKTKTVILTKQNDSTFVOSITKELOANIKYKDAQFONIFQYIG 300
QY 310 GGDSSSSSSSTSNE 322
    |||||
Db 301 GGDSSSSSSSTSNE 313
```

RESULT 3
E86840
Protein maturation protein [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: E86840
R: Bolotin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarne, K.; Welssenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A: Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A: Reference number: A86625; MUID:21235186; PMID:11337471
A: Accession: E86840
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-308 <STO>
A: Cross-references: GB:AE005176; PID:q12724744; PIDN:AAK05823.1; GSPDB:GN00146
A: Experimental source: strain IL1403
C: Genetics:
A: Gene: pmpa

Query Match 32.2%; Score 512; DB 2; Length 308;
Best Local Similarity 36.2%; Pred. No. 3.5e-20;
Matches 114; Conservative 65; Mismatches 120; Indels 16; Gaps 5;
QY 10 MKKLLAGATLLSVATLAACSGSEGADLISMKGDIYTHQFEQVKNPS--A 62

```
Db 1 MKKLLAGATLLSVATLAACSGSEGADLISMKGDIYTHQFEQVKNPSAQOVSQPT 60
    |||||
QY 63 QOVLNNTIQKFEKQYGESELDKEVDVDTIAEKKQYGENYQVRLSQAGMTLETRKAOIR 122
    |||||
Db 61 NTLQNLNTPKIFKIDGKEVDVDTIAEKKQYGENYQVRLSQAGMTLETRKAOIR 120
QY 123 TSRLVELAV-KKVAEALTDKAEKKAFFDEYTPDVTAOIIRLNNEKAEVLEKAKAEAD 181
    |||||
Db 121 TOMEDQADIDHEIKETQYTANLTKAEESHPVDTAVVSETSKDAAATKALDAKRDAG 180
QY 182 PAQLAKNDSTDEKTEKNGEITFDSASTVEPEQVKAFAFDVGVSVITATGQAVSQ 241
    |||||
Db 181 KASPEKTNAE-----SKVTENSTSTVPEVQRAAFKLNKGEESDYESTSSSTGAT 232
QY 242 QYIVKLTKEKSSNIDYKKEKTKTVILTKQNDSTFVOSITKELOANIKYKDAQO 301
    |||||
Db 233 SYIVEMVKTSEKGTDMNKYKELQVYIKTEREQDTTFVSGVIAKYLKKNNTVVKESAF 292
QY 302 NIFTQYIGGDSSSS 316
    |||||
Db 293 STFSQFTQYSSSSSS 307
```

RESULT 4
AF1722
post-translocation molecular chaperone homolog lin2322 [imported] - Listeria innocua
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1722
R: Glaser, P.; Frangoul, L.; Buchtleiser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fshl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A: Authors: Kreif, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A: Title: Comparative genomics of Listeria species.
A: Reference number: AB1077; MUID:21537279; PMID:11679669
A: Accession: AF1722
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-291 <GLA>
A: Cross-references: GB:AL92022; PIDN:CAK97550.1; PID:q16414834; GSPDB:GN00178
A: Experimental source: strain C11P11262
C: Genetics:
A: Gene: lin2322

Query Match 25.2%; Score 400.5; DB 2; Length 291;
Best Local Similarity 33.6%; Pred. No. 2.3e-14;
Matches 103; Conservative 76; Mismatches 111; Indels 17; Gaps 11;

```
QY 10 MKKLLAGATLLSVATLAACSGSEGADLISMKGDIYTHQFEQVKNPSAQOVLNM 69
    |||||
Db 1 MKKLLILGLVMALFSLAAC--GGGNNVTKDSDV-TODELYAMMDKXGSEV--QQL 56
QY 70 TIQKFEKQYGESELDKEVDVDTIAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVEL 129
    |||||
Db 57 TFEKTLGDY--KVSDEVDDKFKKQYQGFSAVLAOSGLTEKSKFSLKYLVLQK 114
QY 130 AVKVAEALTDKAEKKAFFDEYTPDVTAOIIRLNNEKAEVLEKAKAEADPAQLAKDN 189
    |||||
Db 115 ATE--ANMDSKALAEKYYKTMQPDITVSHLVADENKAEVQKLG--DGAFFDLAKEX 171
QY 190 STDKTEKNGEITFDSASTVEPEQVKAFAFDVGVSVITATGQAVSQDYIYVLT 249
    |||||
Db 172 STDATKNGGOLA--PFGSGKMDPAFEKAAVALNKKG--DISAPVKQ--YHIQMD 225
QY 250 KTEKSSNIDYKKEKTKTVILTKQNDSTFVOSITKELOANIKYKDAQFONIFQYIG 309
    |||||
Db 226 KPAITTTDKKAKVKEAYLSQLTLEN--WQTLKEIKYKANVKEVDKDLKDFKDFD 283
QY 310 GGDSSSS 316
    |||
```

Db 284 SASKSS 290

RESULT 5

AC1352

post-translocation molecular chaperone homolog lmc2219 [imported] - *Listeria monocytogenes*

C:Species: *Listeria monocytogenes*

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AC1352

R:Lasar, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Esht, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltounam, A.; Mak, C.; Schleuter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; NCID:21537279; PMID:11679669

A:Accession: AC1352

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-293 <G1A>

A:Cross-references: GB:NC_003210; PIDN:CAD00297.1; PID:g16411689; GSPDB:GN00177

A:Experimental source: strain EGD-e

A:Genetics:

A:Gene: lmc2219

Query Match 24.8%; Score 393.5; DB 2; Length 293;

Best Local Similarity 32.4%; Pred. No. 5, 5e-14;

Matches 100; Conservative 77; Mismatches 115; Indels 17; Gaps 11;

10 MKKKLAGAITLLSVATLAACSGSEGADLISMKGDTVEHQFEQVKNPSAQVLLNM 69

1 MKKKLILGLVMMALPSLAAC--GGGADYVKTDSGVTFDELIDAMKKQYGEFV-QQL 56

70 TIOKVEKQYGSLEDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKQITRKIVEL 129

57 TEKILGDKY--KVSDEVDKRNKEXKSOYDOFSVLTQSGLTESFKQSLYMLYOK 114

130 AVKVAEALTEDEAYKKAPEDEYTPDTAQTIRLNNEDEKAEVLEKKAEGADPAOLAKDN 189

115 ATE--ANTDTSKTKLKYETWQPDITVSHILVADENKAEVDEKLEK--DEKFRADLAKET 171

190 STDEKTEKENGELTFDSASTVEPEOVKKAFAFALDVGVSDVITATGTOVASSQYIVKLT 249

172 STDTATKNDGQGLA-PFGPKMDPAFEKAYALKNG--DISAPVKTQ-YG--YHIIQMD 225

250 KTEKSSNDDYKREKLTLYILQKNDSTFVQSIIGKELOANIKYKQDAFQNIPTQYIG 309

226 KPATKTFEED-KKAVKASYL-ESQLTENMQKTLKREKADANVKEDKDKDAFKDPDG 283

310 GGDSSSSSS 318

284 SSSSDSSS 292

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Query Match 23.6%; Score 375.5; DB 2; Length 299;

Best Local Similarity 31.7%; Pred. No. 4, 9e-13;

Matches 100; Conservative 62; Mismatches 129; Indels 29; Gaps 9;

10 MKKK---LLAGATLLSVATLAACSGSEGADLISMKGDTVEHQFEQVKNPSAQ 64

1 MKKKMLKVLASTATALL--LSGCSNQDQKAVATYSGKVTESFYEKELKOSPPTKT 58

65 VLNMNTQKFEKQYGSLEDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKQITRTS 124

59 MLNMILYRALNHAYGVSSTKTVDAYDSYKQOYGENFAPLSQNGFSRSRESRTN 118

125 KLVELA---VKVAEALTEDEAYKKAPEDEYTPDTAQTIRLNNEDEKAEVLEKKAEGAD 181

119 FLESEVALKRLKAYSESL-----KAWKTYQPKVTVYHILTSDEDTAKQYISDL-AAGKD 172

182 PAOLAKDNSTDEKTEKENGELTFDSASTVEPEOVKKAFAFALDVGVSDVITATGTOVASS 241

173 FAVLAKTDSIDTATKNDGKISFESNNKTIDATFKDAAYKIK-----NGDYTOTPVKV-TN 227

242 QYIIVKLTREKSSNDDYKREKLTLYILQKNDSTFVQSIIGKELOANIKYKQDAFQ 301

228 GYEVKIMNHPAKGT-FTSSKKALTPASVAKMSRDSIMQVISOVLKNOHVITIKDOLA 286

302 NIFTQYIGGGSSSSSTSN 321

287 DAL-----DSYKKPATTN 299

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Db 116 EEFVNEYLPLVLEKLAEGVEITDEEKQAVFENRDSLIEVASHILVEDETAEEVL 175
 QY 173 EKAKAGADFAOLAKONSTDEKTEKNGEITFDSASTVEQVKAFAALDVGVSDVT 232
 Db 176 DRLER-GDFAELAEISVSPSAEANNNDIGFFGKGDVPE-FEEKAFMEIDEVSEPE 233
 QY 233 ATGQAYSSQYIVKLTKEKSSNIDYKEKLTQVILTKONSTFVQSIIGKELQAN 292
 Db 234 ST---YG--YHIIIV---TDRKDSYELEEKIHDTLMNER---SRTQEVRLDLAQN 281
 QY 293 IKVQQAQONIF-----TOYIGGDSSTSSSTSTNE 322
 Db 282 INVLDQFEGFLDLPADAPVEDTPEIDGEDASDAEDQAE 321

RESULT 13

H89970
 hypothetical protein prsa [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: H89970
 R:Kucoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
 ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.;
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; PMID:21311952; PMID:11418146
 A:Accession: H89970
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701634; PIDN:BA042927.1; GSPDB:GN00149
 A:Experimental source: strain N315
 A:Genetic: prsa

Query Match 13.4%; Score 212.5; DB 2; Length 320;
 Best Local Similarity 23.5%; Pred. No. 0.00019;
 Matches 72; Conservative 70; Mismatches 107; Indels 57; Gaps 12;

QY 23 SVATLAACSGSEKAD-----LISMKGDTTEHQFEQVKNPSAQVLLNMTIOKVF 77
 Db 14 SALLIGAC--GASATDSKENTLISSKAGDVTYADIMKKIGKQOIANASTEM-LNKILAD 70
 QY 78 QYSELDDKEDVDITAEKKQY--ENTQVLSQAGMTLETKAQTIRSKLVE--LAVK- 132
 Db 71 KYKNVNDKIKIDEQLEKMQYQYQKDFEKAQOQGLTFADKYKEMLRFAAYHKELSDKI 130
 QY 133 KYAEHELDEA-----YKAFDEYTPDVTAQIIRLNNEKAKAVLEKAKAEGAD 181
 Db 131 KISDEIKEDSKKASHIILIKVKSDEKGLDDKEA-----KQKAEIOREVSNDPEK 183
 QY 182 FQALAKNDSTDEKTEKNGEITFDSASTVEQVKAFAALDVGVSDVTATGQAYSS 241
 Db 184 FGEIKAKESMDGSAKKKDELGY-VLKGOTDKDFEKAFLKLDGEVSEVYKS-----SF 236
 QY 242 QYIYVLTKEKSSNIDYKEKLTQVILTKONSTFVQSIIGKELQANITKVDQAQ 301
 Db 237 GYHIIRKAKPTDFNSEKOSLKEKL-----VDQKVO-KNPKLLLDAYK 277
 QY 302 NIFTOY 307
 Db 278 DLAKY 283

RESULT 14

D97295
 parvulin-like peptidyl-prolyl isomerase, PRSA B. subtilis ortholog [imported] - Clostridium

C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: D97295
 R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: D97295
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-353 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK81151.1; PID:g15026287; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3215

Query Match 11.9%; Score 189.5; DB 2; Length 353;
 Best Local Similarity 26.4%; Pred. No. 0.0035;
 Matches 93; Conservative 59; Mismatches 113; Indels 87; Gaps 21;

QY 6 KGVDAKK-----KLLAGAITLISVATLAACSGSGADLISMK-----GDVITTEHQF 52
 Db 17 RGIRKMSAKQIATALLVGMFTFSAV---GCSVKEKRPAINSKVAVATYGNQITRGER 72
 QY 53 -----YEOVKS-----NPSA-----QOVLNMTIOKPEKQYSGSELDK--- 86
 Db 73 DKLAKGVYQLKSOYQSDYERKNEAVAAALKQKQELTSLIDQKITLKR---AKDQKITL 129
 QY 87 -----EVDITAEKKQY--GENYQVLSQAGMTLETKAQIR---TSKLVELAVK 133
 Db 130 TDEIKTNVDVYAQYQEFTESEKFSOLSQYGYVAEFKPOLKNRAISNKLIOQVVD 189
 QY 134 VAHELITDEAYKKAPEYTPDVT-----AOIIRLNNEKAKAVLEKAKAEGADFAQ 184
 Db 190 V--KVSDEAKKYYDSHNSTQSPNTVHLHL-VKTEKEKAV--KARLDKGEDEAT 243
 QY 185 LAKDNSTDEKTEKNGEITFDSASTVEQVKAFAALDVGVSDVTATGQAYSSQ 242
 Db 244 VAKQYSTD-GSKKEGGDLDIDQENDSNYDKTFMAALKLINDQVSAFVTFQF----- 295
 QY 243 YIIVKLTKEF-KSSNIDYKEKLTQVILTKONSTFVQSIIGKELQANI 293
 Db 296 WHVIRKIKKTEYPAVDKFNVSVDIKQITVLTRO--KSVYQKTLKWKESQANI 345

RESULT 15

E75563
 Probable peptidyl-prolyl cis-trans isomerase, Ppic family - Deinococcus radiodurans (s

C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: E75563
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; PMID:20036896; PMID:10567266
 A:Accession: E75563
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-677 <WHI>
 A:Cross-references: GB:AE001870; GB:AE000513; NID:g6457732; PIDN:AAF09662.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0071
 A:Map position: 1

Query Match 11.8%; Score 187; DB 2; Length 677;
 Best Local Similarity 26.0%; Pred. No. 0.01;
 Matches 91; Conservative 52; Mismatches 125; Indels 82; Gaps 16;

QY 11 KKKLLAGAITLISVATLAACSGSGADLISMKGDTTEHQFEQVKN 59
 Db 27 KRTAVNGILIGISLVYAGMAYQFPNVSGLPFRNGTQFVIVNGAVTAEDLESARRAS 86
 QY 60 P-----SAQVLLNMTIOKVFPEKQYSGSELDKEVD--PTIAEKKQYGEN----- 102

Fri Sep 5 09:24:51 2003

Fri Sep 5 09:24:51 2003

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:21:29 ; Search time 23 Seconds

(without alignments)
658.374 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588

Sequence: 1 SNTYLKGYDMKKKLAGAIT.....IFTYIGGGSSSSSSSTSNE 322

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	375.5	23.6	299 1	PRTM_LACPA
2	367	23.1	299 1	PRTM_LACLC
3	357.5	22.5	292 1	PRSA_BACSU
4	171	10.8	622 1	PPID_HAEIN
5	164.5	10.4	623 1	PPID_ECOLI
6	156.5	9.9	297 1	YACD_BACSU
7	155.5	9.8	282 1	PLP_RICPR
8	151	9.5	282 1	PLP_RICCN
9	146	9.2	273 1	CBP2_CAMJE
10	146	9.2	621 1	PPID_BUCAP
11	132.5	8.3	313 1	SUPA_HAEIN
12	131.5	8.3	623 1	PPID_BUCAT
13	129.5	8.2	1609 1	LMG1_HUMAN
14	127.5	8.0	673 1	PKNX_STRCO
15	127.5	8.0	1790 1	USOI_YEAST
16	126.5	8.0	2116 1	MYSD_DICDI
17	123.5	7.8	544 1	THS_MERTL
18	122.5	7.7	310 1	MFSM_STRGC
19	121.5	7.7	845 1	MYSA_RAT
20	120.5	7.6	555 1	LEB8_DAVXA
21	120	7.6	708 1	RPSD_MYXCA
22	120	7.6	1433 1	RESE_CHICK
23	120	7.6	1959 1	MYH9_CHICK
24	119	7.5	1978 1	MYH9_CHICK
25	118.5	7.5	848 1	NFM_MOUSE
26	118.5	7.5	1841 1	CC12_SCHPO
27	118	7.4	1290 1	RA50_SCHPO
28	118	7.4	1535 1	LM1L_CAEEL
29	118	7.4	1961 1	MYH9_RAT
30	117	7.4	843 1	YMSI_YEAST
31	116.5	7.3	1528 1	SPG2_STRGO
32	116.5	7.3	1528 1	SPG2_STRGO
33	116.5	7.3	1875 1	MLP1_YEAST

34	116	7.3	1014 1	NEBL_HUMAN	076041 homo sapien
35	116	7.3	1300 1	Y338_MYCPN	P75296 mycoplasma
36	116	7.3	1960 1	MYH9_HUMAN	P35579 homo sapien
37	116	7.3	2469 1	TEGO_HSVSA	Q01056 herpesvirus
38	115.5	7.3	274 1	OS44_BORBU	Q04851 borrelia bu
39	115.5	7.3	679 1	YKR9_YEAST	P34237 saccharomyc
40	115	7.2	498 1	HMMR_RAT	P97779 rattus norv
41	115	7.2	1976 1	MYH9_HUMAN	P35580 homo sapien
42	115	7.2	2659 1	RBP1_PRAVB	Q00798 plasmodium
43	114.5	7.2	778 1	TOP1_CANAL	Q00313 candida alb
44	114	7.2	273 1	OS45_BORBU	Q09089 borrelia bu
45	114	7.2	1637 1	MRSP_STRAU	P80544 staphylococ

ALIGNMENTS

```

RESULT 1
ID      PRTM_LACPA          STANDARD:      PRT:      299 AA.
AC      002473;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Protease maturation protein precursor.
GN      PRTM.
OS      Lactobacillus paracasei.
OC      Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC      Lactobacillus.
OX      NCBI_Taxid=1597;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SRRAIN-NCDO 151;
RX      MEDLINE=92381481; PubMed=1512565;
RA      Holck A., Naes H.;
RT      "Cloning, sequencing and expression of the gene encoding the cell-
RT      envelope-associated proteinase from Lactobacillus paracasei subsp.
RT      paracasei NCDO 151."
RL      J Gen Microbiol 138:1353-1364(1992).
CC      -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR PRODUCTION OF ACTIVE FORMS
CC      OF THE SERINE PROTEINASE LOCATED IN OR SECRETED FROM THE CELL
CC      ENVELOPE.
CC      -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC      (Probable).
CC      -1- SIMILARITY: BELONGS TO THE PRTM/PARVULIN FAMILY OF ROTMASES.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: M83946; AAA25247.1; -
CC      PIR: A44858; A44858.
DR      InterPro: IPR000297; Rotomase.
DR      Pfam: PF00659; Rotomase; 1.
DR      PROSITE: PS01096; PPTC_PPIASE_1; 1.
DR      PROSITE: PS01098; PPTC_PPIASE_2; 1.
DR      PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW      isomerase; Rotomase; Membrane; Lipoprotein; signal.
FT      SIGNAL          1..23
FT      CHAIN           24..299
FT      LIPID           24..24
FT      DOMAIN          144..236
FT      DOMAIN          236..236
SQ      SEQUENCE      299 AA; 33145 MW; B9CFIEAB2F8C5867 CRC64;
Query Match      23.68; Score 375.5; DB 1; Length 299;
Best Local Similarity 31.28; Pred. No. 1.9e-13;
Matches 100; Conservative 62; Mismatches 129; Indels 29; Gaps 9;
10 MKKK-----LLAGAILTLAVTLAAGSKSGADLIAMKGDVTEHQFYEQVSNPFAAQ 64

```

```

Db      1 MKKKMKLVLLASTATALL--LSGCOSNQADQKATYSGGKVTESNFKELKQSPPTKT 58
OY      65 VLNMNTIOKVEKQYSELDKEDVDITAEKKQYGENYORVLSQAGMTLETBRAQIRTS 124
Db      59 MLANMLIYRALNHAYGSVSTKTVNDAYDSYKQOYGENFDFLSQNGFSRSFESLRTN 118
OY      125 KLVELA---VKVAEELDEAYKKAFFDEYTPDTVAQIIRLNNEDEKAEVLEKKAEGAD 181
Db      119 FLSEVALKRLKLVSESQL-----KAAMKTYQPKVTYQHILTSDEDTAKQVLSDL-AAKGD 172
OY      182 FALAKNSTDEKTEKNGEITPDSASTVEPQVKKAFALDVGDVDTATNGQAVSS 241
Db      173 FATLAKTDSIDTAKDKNGKISFESNKKTLDTATPKDAAYKLK-----NGDYTOTPVKV-TN 227
OY      242 QYIYVLTKEKTKSSNIDYEKELKTVILTKONQDSTFVOSIIGKELQANIKVKQDAFQ 301
Db      228 GVEYIKMINHPAKGT-FTSSKALTAAYAKWSRDSIMQVLSQVLKNQHVITIKDILA 286
OY      302 NIFQYIGGDSSTSSSTSN 321
Db      287 DAL-----DSYKKKPATYN 299

```

RESULT 2

```

ID      PRM1_LACLC      STANDARD;      PRT;      299 AA.
AC      P14308; P15294;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protease maturation protein precursor.
GN      PRM1.
OS      Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX      NCBI_TaxID=1359;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-SK1; PLASMID-PSK11L;
RX      MEDLINE=89213972; PubMed=2496115;
RA      Vos P., van Asseldonk M., van Jeveren F., Slezén R., Simons G.,
de Vos W.M.;
RT      "A maturation protein is essential for production of active forms of
Lactococcus lactis SK1 serine proteinase located in or secreted from
the cell envelope."
RL      J. Bacteriol. 171:2795-2802(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-WG2;
RX      MEDLINE=89213971; PubMed=2708318;
RA      Haandrikman A.J., Kok J., Laan H., Soemilro S., Ledebor A.M.,
Koudals W.N., Venema G.;
RT      "Identification of a gene required for maturation of an extracellular
lactococcal serine proteinase."
RL      J. Bacteriol. 171:2789-2794(1989).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-MCDO 763; PLASMID-PLP763;
RA      Kivaki M., Ikemura H., Shimizu Kadota M., Hirashima A.;
RL      Submitted (JAN-1989) to the EMBL/GenBank/DBD databases.
RN      [4]
RP      SEQUENCE OF 294-299 FROM N.A.
RC      STRAIN-SK1, and WG2; PLASMID-PSK11L, and pWV05;
RX      MEDLINE=90343335; PubMed=2166472;
RA      Haandrikman A.J., van Leeuwen C., Kok J., Vos P., de Vos W.M.,
Venema G.;
RT      "Insertion elements on lactococcal proteinase plasmids."
RL      Appl. Environ. Microbiol. 56:1890-1896(1990).
CC      -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR PRODUCTION OF ACTIVE FORMS
OF THE SERINE PROTEINASE LOCATED IN OR SECRETED FROM THE CELL
ENVELOPE. PROBABLE PRIASE.
CC      -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

```

```

CC      (Probable).
CC      -1- SIMILARITY: BELONGS TO THE PIC/PARVULIN FAMILY OF ROTAMASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J04962; AAA03532.1; -
DR      EMBL; M26694; AAA60395.1; -
DR      EMBL; X14130; CA132349.1; -
DR      EMBL; M37395; AAA25209.1; -
DR      PIR; A32314; A32313.
DR      PIR; S08083; S08083.
DR      InterPro; IPR000297; Rotamase.
DR      Pfam; PF00639; Rotamase; 1.
DR      PROSITE; PS01096; PIC_PRIASE_1; 1.
DR      PROSITE; PS50198; PIC_PRIASE_2; 1.
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM      Isomerase; Rotamase; Membrane; Lipoprotein; Signal; Plasmid.
FT      SIGNAL 1 23
FT      CHAIN 1 299
FT      LIPID 24 24
FT      DOMAIN 144 236
FT      CONFLICT 46 46
SO      SEQUENCE 299 AA; 3313 MW; 3A7D35DE330ED5B CRC64;

Query Match 23.1%; Score 367; DB 1; Length 299;
Best Local Similarity 31.4%; Pred. No. 5, 4e-13;
Matches 96; Conservative 59; Mismatches 129; Indels 22; Gaps 8;

OY      10 MKKK-----LLAGATILLSVATIAACSKGSEGLDISMGDVTEHQTFEYKSNPSAQO 64
Db      1 MKKKMKLVLLASTATALL--LSGCOSNQDQVATYSGGKVTESNFKELKQSPPTKT 58
OY      65 VLNMNTIOKVEKQYSELDKEDVDITAEKKQYGENYORVLSQAGMTLETBRAQIRTS 124
Db      59 MLANMLIYRALNHAYGSVSTKTVNDAYDSYKQOYGENFDFLSQNGFSRSFESLRTN 118
OY      125 KLVELA---VKVAEELDEAYKKAFFDEYTPDTVAQIIRLNNEDEKAEVLEKKAEGAD 181
Db      119 FLSEVALKRLKLVSESQL-----KAAMKTYQPKVTYQHILTSDEDTAKQVLSDL-AAKGD 172
OY      182 FALAKNSTDEKTEKNGEITPDSASTVEPQVKKAFALDVGDVDTATNGQAVSS 241
Db      173 FATLAKTDSIDTAKDKNGKISFELNKKTLDTATPKDAAYKLK-----NGDYTOTPVKV-TD 227
OY      242 QYIYVLTKEKTKSSNIDYEKELKTVILTKONQDSTFVOSIIGKELQANIKVKQDAFQ 301
Db      228 GVEYIKMINHPAKGT-FTSSKALTAAYAKWSRDSIMQVLSQVLKNQHVITIKDILA 286
OY      302 NIFQY 307
Db      287 DALDSY 292

RESULT 3
ID      PRSA_BACSU      STANDARD;      PRT;      292 AA.
AC      P24327;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Protein export protein prsa precursor (Ec 5.2.1.8).
GN      PRSA.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.

```

RX STRAIN-168;
 RX MEDLINE-92065824; PubMed-19565302;
 RA Kontinen V.P., Saris P., Sarvas M.;
 RT "A gene (prs) of Bacillus subtilis involved in a novel, late stage
 RT of protein export."
 RL Mol. Microbiol. 5:1273-1283(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RC MEDLINE-98240224; PubMed-9579061;
 RX Nodack M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,
 RA Wedler H., Venema G., Bron S.;
 RT "The 172 kb *prkA*-*rddAB* region from 83 degrees to 97 degrees of the
 RT *Bacillus subtilis* chromosome contains several dysfunctional genes,
 RT the *glyB* marker, many genes encoding transporter proteins, and the
 RT ubiquitous *hit* gene."
 RL Microbiology 144:859-875(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RC MEDLINE-99044033; PubMed-9384377;
 RX Kunst F., Oggessara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero L.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Cernerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita Y., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Gutserpi G., Guy B.J., Haga K., Haleh J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapins A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,
 RA Medina N., Melillo R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa K., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Porvillik S., Prescott A.M.,
 RA Presseau E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche A., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Segitunchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Tareuchil M., Tanakoshi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Terasato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*."
 RL Nature 390:249-256(1997).
 RL -1- FUNCTION: INVOLVED IN A LATE STAGE OF PROTEIN EXPORT. PROBABLE
 CC PLASID.
 CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
 CC (omega-20).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO THE PRIC/PARULIN FAMILY OF ROTAMASES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X57271; CAA040543.1; -
 CC EMBL: Y14077; CAA74418.1; -
 CC EMBL: Z99109; CAB12835.1; -
 CC PIR: S15269; S15269.
 CC HSSP: Q13526; 1PIN.
 CC

DR	Subtilisin; BGI0464; prsa.
DR	InterPro; IPR000297; Rotamase.
DR	Pfam; PF00639; Rotamase; 1.
DR	PROSITE; PS01096; PIC_PPIASE_1; 1.
DR	PROSITE; PS01098; PIC_PPIASE_2; 1.
DR	PROSITE; PS00013; PROKAR_LIPOPHOREIN; 1.
KW	Isomerase; Rotamase; Membrane; Lipoprotein; Signal; Complete proteome.
FT	SIGNAL
FT	CHAIN
FT	LIPID
FT	DOMAIN
SO	SEQUENCE
Query Match	22.5%; Score 357.5; DB 1; Length 292;
Best Local Similarity	33.2%; Pred. No. 1.7e-12;
Matches 103; Conservative 60; Mismatches 126; Indels 21; Gaps 10;	
QY	12 KKLGAATLTLVSATILAACSGSGGADLLSKGGVITHEHQYEQVKNSPQAQVLNMTI 71
DB	2 KRIAIAITATISIALSCSSGSDEKVIAKTADGV-TKGELYTYMKKTAGA-SVLTOLVQ 59
OY	72 QKVEKOYGSLELDKEVDVTAEERKQGYGVYQRYLSOAGTLETFRKAQISTYLVELAV 131
DB	60 EKVLDDKY--KVSKEIDNKIKETKQTGLDGYTALEKQYG--DYLKQVAYELLTKRAA 115
OY	132 KKVAEALTEDEAKFDEPTPDVYTAQIIIRLNEDKAEVLEKAKAEGADPAQLAKDNST 191
DB	116 KD-NIVYTADADIHEYMEGLGKIRASHILLVADKKTAEEVEKLIK-KKERFEDLAKEVST 172
OY	192 DEKRKENGCELTTPDSASTVEPEQYKKAFAALDVDCSD-VITATGTQAIYSQYIYIKLK 250
DB	173 DSSASK-GGDGWAFKREQMDERTFSKAFKLTGTEVSPPVKTQYG-----YHI---K 221
OY	251 KTEKSNIDYDKELKTVILNQKNDSEFYQSIIGKELOANIKYKDPAONIFITQYIGG 310
DB	222 KTERGKTDKKELKEVLEEQKINDNAANOEAQYWKAKADIEKDLKDTNTSSTS 281
OY	311 GDSSSSSSSTS 320
DB	282 NSTSSSSSNS 291
RESULT 4	
PPID_HAEIN	STANDARD; PRT; 622 AA.
ID	PPID_HAEIN
AC	P44092;
DT	01-NOV-1995 (Rel. 32, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Peptidyl-Prolyl cis-trans Isomerase D (EC 5.2.1.8) (PPIase D)
GN	(Rotamase D).
OS	ppid OR HT1004.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
OX	NCBI_TaxID=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Rd / KW20 / ATCC 51907;
RX	MEDLINE=95350630; PubMed=7542800;
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., Mc Kenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu T.-I., Glöck A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Heblum E., Cotton M.D., Utecherback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C., FINE L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL	Science 269:496-512(1995).
NN	[2]

```

RP CONCEPTUAL TRANSLATION.
RA Bairoch A.:
RL Unpublished observations (May-1998).
CC -1- FUNCTION: PIPIASES ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
CC INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS (BY
CC SIMILARITY)
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
CC (omega-0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PPIID.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 43 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32781; AAC22665.1; ALT_FRAME.
DR HSSP: O9Y237; 1EQ3.
DR TIGR: H11004; -.
DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIIC_PPIASE_1; 1.
DR PROSITE: PS50198; PPIIC_PPIASE_2; 1.
KM Isomerase; Rotamase; Transmembrane; Periplasmic; Complete proteome.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 17 37 POTENTIAL.
FT DOMAIN 38 622 PERIPLASMIC (POTENTIAL).
FT DOMAIN 270 356 PPIIC.
SQ SEQUENCE 622 AA; 69590 MW; EE5900A5381C4B4 CRC64;

Query Match 10.8%; Score 171; DB 1; Length 622;
Best Local Similarity 21.9%; Pred. 171; 0.026;
Matches 92; Conservative 68; Mismatches 118; Indels 142; Gaps 19;

13 KILAGATITLVAATLAACSGK-----SEGADLISMKGDVTEHQF---YDQ----- 55
16 KFTLG---LIASFIVGMSGYLFSSNDYAAKAVNGEVLISQDFLNRNQEFEIRARRG 72
56 ---YKSNPSAQ-----QYLLNMTIQKVEKQYGSLEKDEVDITAE-----EKKQ 98
73 EAFVAGSDSPERYTALRQNIIVNLMDQLRQYKELKLGVSDEMIKRAIVTDPNFQVKG 132
99 YGEN--YQRYVLSQAGMT-----LEMR 117
133 KPDNAVYQRIILOQNNHLSGDIYASILRASLPLEQIQNGVANSFEIYVQYKNSAEVFFQKR 192
118 KAOIQTSLV-ELAVKVAEAEITD--EAYKRAF----- 148
193 LARLATLSLADEMAKQSVSDDEIKTYEANKQSFVQPEQVKQYIDLSADNISRNQVTD 252
149 ----DETPRYTAQI-----IRLNEDAKAYLEAKAEGADFOALANDSTDEKTKEN 198
253 VETAOYQONKQAFMTQHLAHQIFANEDQAKVAYEELQ-KGANFADVAKAKSLDKISGEN 311
199 GGEITFDSASTVEPEQVKAFAFALVDVSDVITANGTQVSSQYIYVLTKEETESSNI 258
312 GGDILGVN-ENELPKAFEDRAALLOVGYQSPINVDG-----NHAIYQER--KAOSL 362
259 DDYKEKLTQVI-----LTOKQNDSTFVQS-IIGKEQAAINIKYKQDAF--QNI 303
363 ENVKQAIADLVKRSLSMESRYFSLERQASDKAFEDSKSLMTAQAQAVKQESDIYSRQNV 422

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase D (EC 5.2.1.8) (PPIase D)
DE (Rotamase D).
GN PPIID OR B0441 OR C0557.
OS Escherichia coli, and
OS Escherichia coli 06.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick K.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kurd O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / WJ110;
RA Hatada E., Ocmori H., Qiao Y., Tsuji M., Fukuda R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O6:HI / CPT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasco D., Buckles E.L., Liou S.-R., Boutin A., Hackett D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [5]
RP CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=98336187; PubMed=9670013;
RA Dartigalongue C., Raina S.;
RT "A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase
RT required for folding of outer membrane proteins in Escherichia
RT coli.";
RL EMBO J. 17:3968-3980(1998).
RN [6]
RP FUNCTION: PIPIASES ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
RN INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS. ITS PREFERENCE
RN AT THE P1 POSITION OF THE PEPTIDE SUBSTRATE IS GLU > LEU > ALA >
RN HIS > VAL > PHE > ILE > GLY > LYS > THR.
RN -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
RN (omega-0).
RN -1- SUBCELLULAR LOCATION: Type II membrane protein.
RN -1- INDUCTION: BY heat shock.
RN -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
RN STRONG, TO H.INFLUENZAE PPIID.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
DR EMBL: AE000150; AAC7344.1; -.
DR EMBL: U82664; BAB40197.1; -.
DR EMBL: D82943; BAA11645.1; -.

```

DR EMBL: AE016756; AAN79035.1; -
 DR PIR: A64774; A64774.
 DR EcoGene: EG13249; ppID.
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PS01096; PPIC_PPIASE_1; 1.
 DR PROSITE: PS50198; PPIC_PPIASE_2; 1.
 KW Isomerase; Rotamase; Transmembrane; Periplasmic; Heat shock;
 KM Complete proteome.
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 16 36 POTENTIAL.
 FT DOMAIN 37 623 PERIPLASMIC (POTENTIAL).
 FT DOMAIN 266 355 PPIC.
 FT MUTAGEN 312 312 G->A,R: LOSS OF ACTIVITY.
 FT MUTAGEN 313 313 G->A,R: LOSS OF ACTIVITY.
 FT MUTAGEN 347 347 G->A: LOSS OF ACTIVITY.
 FT MUTAGEN 350 350 I->A,F: LOSS OF ACTIVITY.
 SQ SEQUENCE 623 AA: 68150 MW: 0F646F87114A387 CRC64;

Query Match 10.4%; Score 164.5; DB 1; Length 623;
 Best local similarity 22.6%; Pred. No. 0.058;
 Matches 79; Conservative 57; Mismatches 96; Indels 117; Gaps 17;

QY 53 YEQVKS-----PSAQVLLNMTIQKVEKQYSEI-----DKEVDVTI-----AEKKQ 98
 DB 74 YSELANEGYMKTLRQOVNLRLIDEALLDQYARELKLGISDEQVKAIFATPAFYQDKF 133
 QY 99 YGENYRVLSQAGMT-----LETRKAQIRSKIVE----- 128
 DB 134 DNSRYGILNQMGMTDQYQALRNQLTQQLNGVAGDPMKGETDELALVAQORV 193
 QY 129 -----LAVKK-VAEAE-----LTDEAYKKAF----- 148
 DB 194 REATIDVNLAKQPTGEPIEIASYEQNNKNTPEQFQVSYIKLDAATMQPVSADIQ 253
 QY 149 -----DEYT-PDYTA-QIIRLNEDKAKVELEKAGADPAQALAKNSTDEKTEKG 199
 DB 254 SYDDHQDQFQPTQRTYRIIQTETDEKAVYDELIN-KGDPFALAKESADITISARNG 312
 QY 200 GETTFPSASTPEQVQKKAFAFDVGVSDVITATQTAQSSSYIVKLT-KTESSNI 258
 DB 313 GDMGWEDAT-IPDELKNA-----GLKEKGQSGVIRKSSVGFELVRDIDQIPAKVKS 364
 QY 259 DYEKELKTVI-----LTQROND--STFVOSIIGKELOANIK 294
 DB 365 DEVRDDIAKVKHEKALDAVYALQGVKSDAASNDTSLGAE-QAAGVK 412

RESULT 6
 YACD_BACSU STANDARD; PRT; 297 AA.
 AC P37566;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yacd.
 GN YACD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus
 RL DNA Res. 1:1-14 (1994)."
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borisier R., Boursier L., Brans A., Brun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Caprano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Emtlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portelle D., Portwoll S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassaretto A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RL Nature 390:249-256(1997)."
 CC -1 SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC in the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC CC -----

Query Match 9.9%; Score 156.5; DB 1; Length 297;
 Best local similarity 22.5%; Pred. No. 0.067;
 Matches 64; Conservative 68; Mismatches 127; Indels 25; Gaps 10;

QY 14 LLAGATTLISVA-----TLAACKSGSEGDLSMKGDVTEHQFYQVKS--NPSAQVYL 66
 DB 9 IILGALLVCCIAVAAYLTLSQAGASSSGESIAITGKSVTRRWLKEMDQYKSTLEDM 68
 QY 67 LNMTIQVEKQYSGSELDKEDVDITAEKKQYGVQVLSQAGMTLETRKAQIRTSKL 126
 DB 69 INRVVEQLAKKKKLKLSKSEVDRELLIKAVNSRYE-----DEHTEKEMQDIQVNL 124
 QY 127 VELAVKK---VAEELTDEAYKKAFDEYPPDYTAQI--IRLNEDKAKVELEKAKAGAD 181
 DB 125 LEDLITRIDDISNKEI--ESFYNNKKELYOFDDSYRIRHIVYKDEEAREVLKELK-GSS 182
 QY 182 FNLAKDNGSTDEKTEKNGEITF-DSASTPEVQVKKAFALDVGVSDVITATGTAQV 240
 DB 183 FEAVAAERSTDRYTSYGGDLGFTVETASDNIPSAVYTEAKTLKEDMSOEPKIV-----S 237


```

Db      2 KKL---SVLFSLVSMGLAFADKRVATFYKGEVKESSQIMKFRQNLQSGETIKNF 58
QY      60 -----PSAQVLMNTQKVFKEKQYSELDDKVEDDTIAEKKQYGENYQVLSQAGMTLE 115
Db      59 DDEPPDDQKRLIKIYNNLLKE-----EVAKSNITSSKEFOK-----LE 99
QY      116 TRKAQIRTSKIVELAVKVAEALDEAVKKAPEDEYTPDVA--QI-----IRLNEDKAK 169
Db      100 NAKNOALQOELLANYIK-----SNITDKMFDDEYNKRVGNLKGKQELKVAHIILVKSQKEAN 155
QY      170 EYLEKAKAGADPAOLAKDNSTDEKTEKNGEITF---DSASTEVEQYKKAFAALDVG 226
Db      156 DIKTKL-SKGNFTKLAEEISLDKASANGVIGYILNQPGLVPEFEKK-AFALKVNE 213
QY      227 VS-DVTATGTQVASSQYIVKTKTEKSSNIDQKELKYIILQKNDSTFVQSIIG 285
Db      214 VSTPVKTRFG-----WHITIKVLEK-----KPVPIPTKEAKVTIDNTILA 252
QY      286 KE-----LQA-ANIKV 295
Db      253 AEVLKRYIADLEKAKNLKI 271

```

RESULT 9

CBF2_CAMJ2 STANDARD: PRT; 273 AA.

```

AC      046105;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Cell binding factor 2 precursor (Major antigen pcd4a).
GN      CBF2 OR PCD4 OR C10596.
OS      Campylobacter jejuni.
OC      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC      Campylobacteraceae; Campylobacter.
OX      NCBI_TaxID=197;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NCTC 11168;
RX      MEDLINE=20150912; PubMed=10688204;
RA      Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA      Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA      Taggart K., Karlyshev A.V., Moule S., Pallen M.J., Peam C.W.,
RA      Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA      Whitehead S., Barrell B.G.;
RT      "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT      reveals hypervariable sequences.";
RL      Nature 403:665-668(2000)
CC      -1- SIMILARITY: BELONGS TO THE PIC/PAVULIN FAMILY OF ROTAMASES.
CC      -1- STRONG, TO H. PYLORI HP0175.

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

DR      EMBL: X84703; CA59175.1; -
DR      EMBL: AL139075; CAB75232.1; -
DR      PIR: S52412; S52412.
DR      HSSP: Q13526; 1PIN.

```

```

DR      InterPro: IPR000297; Rotamase.
DR      Pfam: PF00639; Rotamase; 1.
DR      PROSITE: PS01096; PPI_CPIASE_1; 1.
DR      PROSITE: PS0198; PPI_CPIASE_2; 1.
KW      Isomerase; Rotamase; Antigen; Signal; Complete proteome.
FT      CHAIN          1      21      SIGNAL.
FT      DOMAIN         22      273      CELL BINDING FACTOR 2.
FT      DOMAIN         131     228      PPI_C.
SQ      SEQUENCE      273 AA; 30518 MW; 1C014658BBD439 CRC64;

```

Query Match 9.2%; Score 146; DB 1; Length 273;
 Best Local Similarity 24.8%; Pred. No. 0.22;
 Matches 76; Conservative 47; Mismatches 101; Indels 82; Gaps 14;

```

QY      10 MKR-----KLGAATLLSVATLACSGSSEADLSMKGVTE---HOPY-----E 54
Db      1 MKRFSVAATLNGVLANNAVA-----TNGSISISTEVESEFPAPLRQ 48
QY      55 QVKSNSPAQVLMNTQKVFKEKQYSELDDKVEDDTIAEKKQYGEN---YQVLSQAG 111
Db      49 DFKTLPDNQ-----KKALIQYIMQ-----DLIDAKKQNLKRPVITKELDR- 93
QY      112 MLETRKAOIRTSKIVELAVKVAEALDEAVKKAPEDEYTPD-----VTAQITRLNN 164
Db      94 -----KDAILVNVQEKILNTIKIDAKVKAFYQDNKDKYKPARVOAKHILVAT 143
QY      165 EDKAKEVLE-----KKAAGADPAOLAKDNSTDEKTEKNGEITFDSASTEVEQYKKA 219
Db      144 EKKAQIINELKGLKRELDAKFSLEAKRSIDPSGSKNGELGWPDSTW- KPPTDA 202
QY      220 FALDVGVSVDVITATGTQVASSQYIVKTKTEKSSNIDYK-----EKLTVIL 270
Db      203 FALK-----NGTITTTVKINFGYHVLKENSQAKGIRKDEVKQGIENLKFEEFKYI- 257
QY      271 TOKQND 276
Db      258 NQKQGD 263

```

RESULT 10

PPID_BUCAP STANDARD: PRT; 621 AA.

```

AC      08K987;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Peptidyl-prolyl cis-trans isomerase D (EC 5.2.1.8) (PPIase D)
DE      (Rotamase D).
GN      PPID OR BUSG462.
OS      Buchnera aphidicola (subsp. Schizaphis graminum).
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Buchnera.
OX      NCBI_TaxID=98794;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22084549; PubMed=12089438;
RA      Tamas I., Klasson I., Canback B., Naeslund A.K., Eriksson A.-S.,
RA      Werngreen J.U., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RT      "50 million years of genomic stasis in endosymbiotic bacteria.";
RL      Science 296:2376-2379(2002).
CC      -1- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS (BY
CC      SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
CC      (omega-0).
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC      -1- SIMILARITY: BELONGS TO THE PIC/PAVULIN FAMILY OF ROTAMASES.
CC      -1- STRONG, TO E. COLI PPID.

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE014122; AAM68005.1; -

DR InterPro; IPR000297; Rotamase.

DR Pfam; PF00639; Rotamase; 1.

DR PROSITE; PS01096; PPTC_PP1ASE_1; FALSE-NEG.

DR PROSITE; PS50198; PPTC_PP1ASE_2; 1.

KW Isomerase; Rotamase; Transmembrane; Complete proteome.

FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 11 31 POTENTIAL.

FT DOMAIN 32 621 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 228 355 PPTC.

SO SEQUENCE 621 AA; 74139 MW; 5B7A89B253144C7C CRC64;

Query Match 9.2%; Score 146; DB 1; Length 621;

Best Local Similarity 25.1%; Pred. No. 0.55;

Matches 66; Conservative 51; Mismatches 80; Indels 66; Gaps 15;

QY 56 VKSNPSAOVLLNMTQKVEKQYSELDKEDVDITAE-KKQYGENYQVLSQAGMTL 114

DB IKINP-----IINN--OKYTEKEINNYFDQHKNEFYEPKFKISY-----IQI 238

QY 115 ETRKAQIRTSKVELAVKVAEALDEAYKKAFFDEYTPVTAQ--IIRLNEDKAKEVL 172

DB KPKKFKIQCSN-----EETKNWTKKNIDKITSNOERQYSIITQTKTNEALSLI 286

QY 239 KPKKFKIQCSN-----EETKNWTKKNIDKITSNOERQYSIITQTKTNEALSLI 286

QY 173 EKAKAGADFAQAKDNSTDEKTEKNGEITPDASTPEQYKKAFAALDVGVSDVIT 232

DB 287 SELK-KGEPSFKIAKEKSIDPFSSEOGNGITKNF-VNDEIKIANLE-KIDQISNIT- 342

QY 233 ATGTQAVSSQYVYVLT-----KTEKSSNID-----YKELKT-----VITQKQ 274

DB 343 -----KFNNEFLIKLKLIPKPKYKISEVSDLIENEIKYKSLNTYKRLKDKIAITAIK 397

QY 275 NDSTFVOSIIGKELQANIKVKD 297

DB 398 NINRF--DLI--LKKTNILPKK 415

RESULT 11

SURF_HAELN STANDARD; PRT; 313 AA.

AC P44721;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

GN SURF OR HI0458.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-Rd / KM20 / ATCC 51907;

RA MEDLINE=95350630; PubMed=7542800;

RA Fleischnann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kervatage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Gocayne J.D., Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Kelley J.M., Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M., Welman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.,

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."

RL Science 269:496-512(1995).

CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC -1- SIMILARITY: BELONGS TO THE PPTC/PARVULIN FAMILY OF ROTAMASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U32728; AAC22116.1; -

DR PIR; P64069; P64069.

DR HSSP; Q91237; LFJD.

DR TIGR; H10458; -

DR InterPro; IPR000297; Rotamase.

DR Pfam; PF00639; Rotamase; 1.

DR PROSITE; PS01096; PPTC_PP1ASE_1; 1.

DR PROSITE; PS50198; PPTC_PP1ASE_2; 1.

KW Isomerase; Rotamase; Signal; Complete proteome.

FT SIGNAL 1 27 POTENTIAL.

FT CHAIN 28 313 SURVIVAL PROTEIN SURF HOMOLOG.

FT DOMAIN 167 267 PPTC.

SO SEQUENCE 313 AA; 34669 MW; EAE2A7C352302450 CRC64;

Query Match 8.3%; Score 132.5; DB 1; Length 313;

Best Local Similarity 22.4%; Pred. No. 1.3;

Matches 70; Conservative 62; Mismatches 110; Indels 71; Gaps 15;

QY 12 KKLGAATITLAVATLAACSKG-SEGADLSKKGIVTGHQYEOVKSN-----PSA 62

DB 4 KKFVLSFLFATLALGCAFTSMQAEERVAATVDGIVLE-----SOVRAMKGGKGRQSI 59

QY 63 QOVLNMTIQKVEKQYSELDKEDVD--DTIAEEKQYGENYQVLSQAGMTLETRK 118

DB 60 DTIIDILQKAVQES-GVKIDPREIDHIVEDTJAARNGLTGQ-FLDALDYGISLNTFR 117

QY 119 AQIRT-----SKIVELAVKVAEALDEAYKKAFFDEYTPVTAQ.158

DB 118 QOIANOMWGAIVKRAIQESIDVTREYVALQKMLDEKSGCTQKVTGKEY--EVHRI 175

QY 159 IIRLN--NEDRAKEVLEKAKAE--GADFAQAKDNSTDEKTEKNGEITPDASTEV 211

DB 176 LKTLNPLNDQAQKQAKIRSDIAGKTTFADALAKYSKDYLSGANGSGIYAPPEYIA 235

QY 212 PEQVK-----KAFA--LDVGVSD-VITANGTQAVSSQYIYVLTQKTE 253

DB 236 PQFAQTVVSKSGVTSAPKTEFGWHIIEVGVGRDGLTA--EAYTQKAEVRLVNTLOQ 292

QY 254 KSSNIDYKELK 266

DB 293 DATN--DWYKALR 303

RESULT 12

PPID_BUCAI STANDARD; PRT; 623 AA.

AC P57550;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Peplidy1-Proly1 cis-trans isomerase D (EC 5.2.1.8) (PPIase D) (Rotamase D).

GN PPID OR B0478.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI_TaxID=118099;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-Tokyo 1998;

RA MEDLINE=20445173; PubMed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS."

RL Nature 407:81-86(2000).

```

CC -1- FUNCTION: PEPIASES ACCELERATE THE FOLDING OF PROTEINS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PPID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001119; BAB13175.1; -.
DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00659; Rotamase; 1.
DR PROSITE: PS01096; PPIC_PPIASE_1; FALSE_NEG.
DR PROSITE: PS50198; PPIC_PPIASE_2; 1.
KW Isomerase; Rotamase; Transmembrane; Complete proteome.
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 12 32 POTENTIAL.
FT DOMAIN 33 623 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 267 357 PPIC.
SQ SEQUENCE 623 AA; 74261 MW; 3CF307543EBE01F6 CRC64;

Query Match 8.3%; Score 131.5; DB 1; Length 623;
Best Local Similarity 19.1%; Pred. No. 3.2;
Matches 65; Conservative 62; Mismatches 92; Indels 121; Gaps 13;

QY 66 LAMTQKVEKQYSELDDEKVDYIAE----- 94
DB 91 VLSGLNNVLEEGYAKAMNNTLVNDNTIKTIYNSPIFOKNKSKERYLVLYLSINSTN 150
QY 95 -----EKQYQENTQRYVLSQAGWLETKRAQI---RTSKVELAVKRV----- 134
DB 151 HEYINIKKKIKNTINELHTITSKSNFLLKKEKKIKLLSOKRIKKAIVDPSIYKNI 210
QY 135 -----AEAEI-----TDEAY-----KKAFDETPVTA----- 157
DB 211 TNOEAQYFKKKNQDNFYPEKFINVELKTDNFKHCENKEIYDWTIRNTQYSTKEKR 270
QY 158 --OIRINDEKAKVELEKAKAGADPAOLAKNSTDEKTEKNGEITFSPASVEPQV 215
DB 271 RYSLIOYKKNQQAISILSRHLNTPEDFSKTAQEGSTDPISKKGDIQWISIDL-IPDEI 329
QY 216 KKAAPALDVGVSDVITATGTQAYSSQYIVKLT-----KTEKSSN 257
DB 330 KHAHLN-KKNQISDVI-----PFHNEFLIVKLNETQIGTQKRIYEVDSIKQIKQKS 382
QY 258 IDDYKKEKLVILTKQNDSTFVOSITGKELQANIKVD 297
DB 383 LDLYNE-LKKNISNKLNDPKIERI---LKNNNIIIOE 417

RESULT 13
LMG1_HUMAN STANDARD; PRT; 1609 AA.
AC P11047;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Laminin gamma-1 chain precursor (Laminin B2 chain).
GN LAMC1 OR LAMB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093128; PubMed=1985895;

```

```

RA kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;
RT "Structure of the human laminin B2 chain gene reveals extensive
RT divergence from the laminin B1 chain gene.";
RL J. Biol. Chem. 266:221-228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198245; PubMed=3360804;
RA Plikarainen T., Kallunki T., Tryggvason K.;
RT "Human laminin B2 chain. Comparison of the complete amino acid
RT sequence with the B1 chain reveals variability in sequence homology
RT between different structural domains.";
RL J. Biol. Chem. 263:6751-6758(1988).
RN [3]
RP SEQUENCE OF 1393-1609 FROM N.A.
RX MEDLINE=89169663; PubMed=3234037;
RA Pukushima Y., Plikarainen T., Kallunki T., Eddy R.L., Byers M.G.,
RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;
RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
RT the gene to chromosome region 1q25-->q31.";
RL Cytogenet. Cell Genet. 48:137-141(1988).
RN [4]
RP SEQUENCE OF 1282-1609 FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=92216123; PubMed=1806043;
RA Santos C.L.S., Sabdaga J., Brentani R.;
RT "Differences in human laminin B2 sequences.";
RL DNA Seq. 1:275-277(1991).
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 1 laminin IV domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55210; AAA59492.1; -.
DR EMBL: M55217; AAA59492.1; JOINED.
DR EMBL: M55201; AAA59492.1; JOINED.
DR EMBL: M55211; AAA59492.1; JOINED.
DR EMBL: M55212; AAA59492.1; JOINED.
DR EMBL: M55213; AAA59492.1; JOINED.
DR EMBL: M55214; AAA59492.1; JOINED.
DR EMBL: M55215; AAA59492.1; JOINED.
DR EMBL: M55216; AAA59492.1; JOINED.
DR EMBL: M55192; AAA59492.1; JOINED.
DR EMBL: M55193; AAA59492.1; JOINED.
DR EMBL: M55194; AAA59492.1; JOINED.
DR EMBL: M55195; AAA59492.1; JOINED.
DR EMBL: M55196; AAA59492.1; JOINED.
DR EMBL: M55197; AAA59492.1; JOINED.
DR EMBL: M55198; AAA59492.1; JOINED.

```

FT	DISULFID	754	770	BY SIMILARITY.
FT	DISULFID	773	781	BY SIMILARITY.
FT	DISULFID	775	792	BY SIMILARITY.
FT	DISULFID	795	804	BY SIMILARITY.
FT	DISULFID	807	825	BY SIMILARITY.
FT	DISULFID	828	842	BY SIMILARITY.
FT	DISULFID	830	849	BY SIMILARITY.
FT	DISULFID	852	861	BY SIMILARITY.
FT	DISULFID	864	881	BY SIMILARITY.
FT	DISULFID	884	898	BY SIMILARITY.
FT	DISULFID	886	905	BY SIMILARITY.
FT	DISULFID	907	916	BY SIMILARITY.
FT	DISULFID	919	932	BY SIMILARITY.
FT	DISULFID	935	947	BY SIMILARITY.
FT	DISULFID	937	954	BY SIMILARITY.
FT	DISULFID	956	965	BY SIMILARITY.
FT	DISULFID	968	980	BY SIMILARITY.
FT	DISULFID	983	995	BY SIMILARITY.
FT	DISULFID	985	1001	BY SIMILARITY.
FT	DISULFID	1003	1012	BY SIMILARITY.
FT	DISULFID	1015	1028	BY SIMILARITY.
FT	DISULFID	1031	1031	INTERCHAIN (PROBABLE).
FT	DISULFID	1034	1034	INTERCHAIN (PROBABLE).
FT	DISULFID	1600	1600	INTERCHAIN (PROBABLE).
FT	CARBOHYD	60	60	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	650	650	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1022	1022	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1107	1107	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1161	1161	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1205	1205	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1223	1223	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1241	1241	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1380	1380	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1395	1395	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1439	1439	N-LINKED (GLCNAC. . .)
FT	VARIANT	458	458	V -> I (IN dBSNP: 20563).
FT				/FT14-VAR_014700.

Query Match	8.2%	Score 129.5	DB 1	Length 1609
Best Local Similarity	20.9%	Pred. No. 12		
Matches 65	Conservative	73	Mismatches 136	Indels 37
				Gaps 13
QY	32 KCESGADLISMKG--DVITEHQFEQVYVNSNAQVLLNMVTIOKVFKEQYGESELDK--	86		
Db	1048 KIQELESLLANGTGDEMTDQAFEDRLK---EAEREVYDLREADVDVQVONLMDRIQ	110		
QY	87 EVDQDTIAEKKQYGEYQVAVLSOAGMTLETKRAQI-RTSKIVELAVKVAEALTFDEAK	145		
Db	1105 RVNNTTSSQISIRL-QNIRNTIEETGTLADQARAHVNTERLLIEISRELEKKVVAANYS	116		
QY	146 KAFDEYTPDVTAQIIRLNN---EDRAKEVLEKAKAEGADFAQOLAK--DNSTDE----	193		
Db	1164 VHQPESTGDP-----NNMTLLAEBAERKLAERHKOEADIVYAVTANDSTEAAYNLL	121		
QY	194 KTKENGELTFD-SASTVEPEQYKKAAPLADVQ--VSDVYATATGTQA--YSSQYIYK	247		
Db	1217 RTLAGENQYAFEEELNKRTEQAKNISODLEKQAAVYHEAKRAGKAYEIVASVQLS	127		
QY	248 LTRKT--EKSSNIDYKELIKTVILTKQONDSTFVOSIIGKELQAAINIKYKQDAFQNIPT	305		
Db	1277 LDSETLENEANNIKMAENLEQLIDQIKDYEDLREDMRKELEVRNLLLEKGTQOOTAD	133		
QY	306 QYIGGDDSSSS	316		
Db	1337 QLLARADAKA	1347		

```

AC Q9XA16;
DR 15-SEP-2003 (Rel. 42, Created)
DR 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable serine/threonine-protein kinase SC03848 (EC 2.7.1.37).
GN SC03848 OR SCH69.18.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Latke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417.141-147(2002).
CC -1- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphoprotein.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 4 PASTA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL939118; CAB45215.1; -
DR PIR: T36717; T36717.
DR InterPro: IPR005543; PASTA.
DR InterPro: IPR000719; Prot_Kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR Pfam: PF03793; PASTA. 4.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00740; PASTA. 4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF; 1.
DR Hypothetical protein: Transferase; Serine/threonine-protein kinase;
KW ATP-binding; Repeat; Complete proteome.
FT DOMAIN 11 277 PROTEIN KINASE.
FT DOMAIN 379 445 PASTA 1.
FT DOMAIN 446 511 PASTA 2.
FT DOMAIN 512 580 PASTA 3.
FT DOMAIN 581 649 PASTA 4.
FT NP_BIND 17 25 ATP (BY SIMILARITY).
FT BINDING 40 40 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 673 AA; 71488 MW; 42A0615E239722DE CRC64;
Query Match 8.0%; Score 127.5; DB 1; Length 673;
Best Local Similarity 22.1%; Pred. No. 5.7;
Matches 70; Conservative 45; Mismatches 151; Indels 51; Gaps 11;
OY 14 LLAGATITLIS---VATLAACSGSEGADLISMKGVITEHOFYEOYKSNPSAQOVLNMT 70
DB 354 VLAGVLYLVGAILIGKFAFGSDGPGNDKVPAPAF-----GLSKADAQQQADNID 404
OY 71 IOKVFEKQYSGELDEVDVDTIAEKKQYKGENQVRLSQAGMTLEFRKQIRTSKIVELA 130
DB 405 LVTFEKKQ---ECEDQPKGNICAQDPKQ-GTDVKE-STVNLVSTGAPKV-----A 451

```

```

OY 131 VKKVAEELTDEAYKKAPEDEYTPDYTAQIIRLNNEKAKEVLEKAKGADPAQLAKDNS 190
DB 452 VENVIDKNI-DEAKKQI-----EDKGEVETKQTQESQDDGCTILSQNP 493
OY 191 TDEKTRKENGELTFPDSASTE---VPEQYKKA--FALDVDSVDPVATATQAVSSOY 244
DB 494 DPKLELEKSTVTLLEVAKAEKATVPDYVGRFCDEAKAQVESGDLTAVCTQDPNIDNPQ 553
OY 245 IYKILTKTRKSSNIDYKREKTLVITLTONDSTFVQSLTGELQAAIKVDAQFQNF 304
DB 554 VGRVISTTPQSSSTOYDPGSKV--TIYGVKAVERTKVEYRGKTLAEARQILOQSGFTNVO 611
OY 305 TOYIGGDDSSSSSTSN 321
DB 612 VAQSGDDNNAVFASN 628
RESULT 15
USOL YEAST STANDARD; PRI: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54378; CAA38253.1; -
DR EMBL: L03188; AAB00143.1; -
DR EMBL: U53668; AAB66559.1; -
DR SGD: S0002216; USOL.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR006955; Usol_p115_C.

```

```

DR InterPro: IPR006953; Usol_p115_head.
DR Pfam: PF04871; Usol_p115_C; 1.
DR Pfam: PF04869; Usol_p115_head; 1.
KM Transport: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DREDD (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 8.0%; Score 127.5; DB 1; Length 1790;
Best Local Similarity 22.8%; Pred. No. 17;
Matches 69; Conservative 59; Mismatches 103; Indels 71; Gaps 15;

QY 49 EHOFEQVK-----SNPSAQVILNMTIQVFEKQYG-----SELDKQVD--D 90
DB 1172 EROYNEIISQLNDEITSTQOE--NESIKRKNDELGEVAKMKTSEOSNLKKSEIDALN 1229
QY 91 TIAEKKOYGENYQRYVLSQAGMTLETRKAOI-----RTSKIVELAVK----- 132
DB 1230 LQTKELKKNENYEALESIKSVESEYTKIKELQDECNFKKEVSELEDKLLKASEDKNS 1289
QY 133 KYAEALTPDAKKAFDEYTPDVTAGIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTD 192
DB 1290 KYLELOKESSEKIKKEILDATTEIKIQLEKITYNLSKAKE-----KSE-SELSRLKKTSSSE 1343
QY 193 EKTRENGEITPDSASTVEPDQYKKAFAFDVGVSDVITATGTQAYSSQY-----YIVK 247
DB 1344 RKNAE-----QLEKLNKI--QIKNAFEREKRLNKG--SSTTQYSEKINTLEDELIR 1396
QY 248 LFKKTE-KSSNIDYKELKTV-----ILTKQNDSTFVOSIIGKELQANIKVQDAF 300
DB 1397 LQNNELKAKEIDNTRSELEKVSLSNDELLEEKQNTI-----KSLQDEILSYKDKIT 1448
QY 301 QN 302
DB 1449 RN 1450

```

Search completed: August 29, 2003, 10:34:25
 Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:30:34 ; Search time 99 Seconds
(without alignments)
839.323 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 158
Sequence: 1 SNTYLKGYDMKKLLAGAIT.....IFTGYIGGDSSTSSSTSE 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	97.0	313	16	097R51 streptococc
2	1538	96.9	313	16	08DQ24 streptococc
3	795	50.1	333	16	08CVC6 streptococc
4	738	46.3	309	16	08E602 streptococc
5	736	46.3	309	16	08E0C6 streptococc
6	696	43.8	351	16	099256 streptococc
7	687	43.3	351	16	08P0E5 streptococc
8	619	39.0	309	16	099X79 streptococc
9	614	38.7	309	16	08K5F3 streptococc
10	512	32.2	308	16	09CEV9 lactococcus
11	400.5	25.2	291	16	0929F4 listeria in
12	393.5	24.8	293	16	08Y557 listeria mo
13	373	23.5	299	2	09A1O3 lactococcus
14	363	22.9	295	2	053962 lactococcus
15	328	20.7	294	16	08Y759 listeria in
16	327.5	20.6	294	16	092BR2 listeria in

17	310	19.5	299	16	08CXK4 oceanobacil
18	286	18.0	333	16	09KDN4 bacillus ha
19	265	16.7	248	2	08GCH6 bacillus ce
20	214.5	13.5	306	16	08R760 thermoaer
21	212.5	13.4	320	16	099T36 staphylococ
22	208	13.1	325	16	08CNR4 staphylococ
23	189.5	11.9	353	16	097E99 clostridium
24	187	11.8	677	16	09RY79 delnoccoccus
25	174.5	11.0	626	16	09CJMO pasteurella
26	168	10.6	619	16	08DG31 vibrio vuln
27	163.5	10.3	623	16	08XE55 escherichia
28	163	10.3	282	16	08XPV2 ralstonia s
29	162	10.2	628	16	08ZC69 yersinia pe
30	158.5	10.0	623	16	08ZRB8 salmonella
31	158.5	10.0	623	16	08ZRU9 salmonella
32	158	9.9	619	16	09KQ10 vibrio chol
33	155.5	9.8	281	16	08XU7 ralstonia s
34	155.5	9.8	336	16	051135 borrelia bu
35	155	9.8	303	16	098G68 rhizobium m
36	154	9.7	621	16	08EG35 shewanella
37	152	9.6	82	2	08YS96 streptococc
38	152	9.6	248	16	08XNR4 clostridium
39	151.5	9.5	342	16	08XHK0 brucella me
40	151	9.5	331	16	08YJG0 rhizobium m
41	150.5	9.5	284	16	092MJ0 brucella su
42	150.5	9.5	331	16	08FYE0 drosophila
43	150.5	9.5	7210	5	09Y7G8 drosophila
44	150.5	9.5	9270	5	08MUD9 drosophila
45	148	9.3	1434	5	081492 plasmodium

ALIGNMENTS

RESULT 1
ID 097R51 PRELIMINARY; PRT; 313 AA.
AC 097R51;
DT 01-OCT-2001 (TREMREL. 18, Created)
DT 01-OCT-2001 (TREMREL. 18, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Protease maturation protein, putative.
GN SP0981.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anglucci S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL: AE007401; ANK45102.1.
DR TIGR: SP0981.
DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS0198; PPIC_PP1ASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 313 AA; 34412 MW; 42FEF736A0ABED0 CRC64;
Query Match 97.0%; Score 1541; DB 16; Length 313;
Best Local Similarity 100.0%; Pred. No. 3.4e-75;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

QY 10 MKKLLAGATLLSVATTLAAGSGEGADLISMKGDVITEHOFTYQVKNPSAQOVLNM 69
DB 1 MKKLLAGATLLSVATTLAAGSGEGADLISMKGDVITEHOFTYQVKNPSAQOVLNM 60
QY 70 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 129
DB 61 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 120
QY 130 AVKVAEALTDPAKAFDEYTPDVTQAIIIRLNEDKAEVLEKAKAGADFAQLAKDN 189
DB 121 AVKVAEALTDPAKAFDEYTPDVTQAIIIRLNEDKAEVLEKAKAGADFAQLAKDN 180
QY 190 STDEKTKENGGEITFDASASTEVPEQVKKAFALDVGVSDVITATGTAQVSSQYIYVLT 249
DB 181 STDEKTKENGGEITFDASASTEVPEQVKKAFALDVGVSDVITATGTAQVSSQYIYVLT 240
QY 250 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANINIKVQDAFONITFYQYG 309
DB 241 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANINIKVQDAFONITFYQYG 300
QY 310 GGDSSSSSSSTSNE 322
DB 301 GGDSSSSSSSTSNE 313

```

RESULT 2

```

ID 08DQ24 PRELIMINARY; PRT; 313 AA.
AC 08DQ24:
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Proteinase maturation protein.
GN PPMa OR SPR0884.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgelt S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Puller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAlister S.M., McInerney M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.I.,
RA Glas J.I.,
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008462; AAK99688.1;
KW Complete proteome.
SQ SEQUENCE 313 AA; 34439 MW; 44063A36ABF9BED0 CRC64;

```

Query Match 96.9%; Score 1538; DB 16; Length 313;
 Best Local Similarity 99.7%; Pred. No. 5e-75;
 Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 MKKLLAGATLLSVATTLAAGSGEGADLISMKGDVITEHOFTYQVKNPSAQOVLNM 69
DB 1 MKKLLAGATLLSVATTLAAGSGEGADLISMKGDVITEHOFTYQVKNPSAQOVLNM 60
QY 70 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 129
DB 61 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 120
QY 130 AVKVAEALTDPAKAFDEYTPDVTQAIIIRLNEDKAEVLEKAKAGADFAQLAKDN 189
DB 121 AVKVAEALTDPAKAFDEYTPDVTQAIIIRLNEDKAEVLEKAKAGADFAQLAKDN 180

```

```

QY 190 STDEKTKENGGEITFDASASTEVPEQVKKAFALDVGVSDVITATGTAQVSSQYIYVLT 249
DB 181 STDEKTKENGGEITFDASASTEVPEQVKKAFALDVGVSDVITATGTAQVSSQYIYVLT 240
QY 250 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANINIKVQDAFONITFYQYG 309
DB 241 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANINIKVQDAFONITFYQYG 300
QY 310 GGDSSSSSSSTSNE 322
DB 301 GGDSSSSSSSTSNE 313

```

RESULT 3

```

ID 08CVC6 PRELIMINARY; PRT; 333 AA.
AC 08CVC6:
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Putative protease maturation protein precursor.
GN PRM OR SMU.648.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdile D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Fetherly J.J.,
RT "Genome sequence of Streptococcus mutans JAI59, a cariogenic dental
RT pathogen."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE014908; AAN58382.1;
KW Protease; Complete proteome.
SQ SEQUENCE 333 AA; 36092 MW; 666C84A5639C54E CRC64;

```

Query Match 50.1%; Score 795; DB 16; Length 333;
 Best Local Similarity 50.0%; Pred. No. 3.1e-35;
 Matches 156; Conservative 66; Mismatches 84; Indels 6; Gaps 2;

```

QY 10 MKKLLA-GATLLSVATTLAAGSGEGADLISMKGDVITEHOFTYQVKNPSAQOVLNM 68
DB 1 MKKRLTATGLVTLSTVTLAAGSKTQNSKIATMGDTITVADFVENVNSTASQAVLS 60
QY 69 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 128
DB 61 TIOKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETAKAQTIRTSKLYEL 120
QY 129 LAVKVAEALTDPAKAFDEYTPDVTQAIIIRLNEDKAEVLEKAKAGADFAQLAKDN 188
DB 121 YAVKEAKKEITDASAKSAVKDYKPEVTAQVQLDSEDAKAVLEPAKADGADFAKIAKD 180
QY 189 NSTDEKTKENGGEITFDASASTEVPEQVKKAFALDVGVSDVITATGTAQVSSQYIYVLT 248
DB 181 NTKGKRT-----EYSPDSGSTNLPQSVLSAALNLKDGVSIVIKASDSTTYAPVYIYKI 235
QY 249 TTKTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANINIKVQDAFONITFYQYG 308
DB 236 TTKTDKNAWMAKAYKKRKLKELIYSOKLUNDSFNMAVIGKAFKANKVKKIDKASEILSOYA 295
QY 309 GGDSSSSSSSTS 320
DB 296 AASGSGSGSTT 307

```

```

RESULT 4
ID 08E602 PRELIMINARY; PRT; 309 AA.
AC 08E602;

```


DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS0827.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=22242308; PubMed=12354221.
RA Glaser P., Rumsloh C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Talloni L., Poyart C., Tlieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease";
RL Mol. Microbiol. 45:1498-1513(2002).
DR EMBL: AL766847: CADD4411.3; .
DR Sagaristat; gbs0827; .
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 309 AA; 33852 MW; 813F3C378D6315A6 CRC64;

Query Match	46.5%;	Score 738;	DB 16;	Length 309;
Best Local Similarity	48.4%;	Pred. No. 3.2e-32;		
Matches 149;	Conservative 62;	Mismatches 89;	Indels 8;	Gaps 3

[illegible]

RESULT 5

ID Q8E0C6 PRELIMINARY; PRT; 309-AA.

DT 01-MAR-2003 (TremblRel. 23, Created)

DT 01-MAR-2003 (Tremblé, 23, Last annotation update),

OS strantccccccu aca]act+aa (serotino v)
GN SAG0808.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus

NCBI_JAXID=210400;	
OX	[1]
RN	

RC STRAIN=2603 V/R /

RA Tettelein H.: Masigmanl V.: Cleslew

RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,

RA Deacy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarcellini M., Mora M.,
RA Jacobini E.T., Brettoni C., Gall G., Mariani M., Vegeni F., Malone D.,
RA Rinaldo D., Rappunoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V *Streptococcus agalactiae*.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL: AE014228; AAC99695.1; -
DR TIGR: SAG0808; -
KW Complete proteome.
SQ SEQUENCE 309 AA; 33890 MW; 3689E218DBD37810 CRC64;

Query Match	46.3%;	Score 736;	DB 16;	Length 309;
Best Local Similarity	48.4%;	Pred. No. 4.1e-32;		
Matches 149;	Conservative 61;	Mismatches 90;	Indels 8;	Gaps 3;

[illegible]

RESULT 6

ID Q99256 PRELIMINARY; PRT; 351 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-OCT-2002 (TREMBlé). 22, Last annotation update)

GN SPY1390 OR PRTM.1 OR SPYM3_1059

Streptococcus pyogenes (serotype M3)

NCBI TaxID=1314. 198466

RP	SEQUENCE FROM N.A.
RN	[11]

AC SPECIES-S: pyogenes; strain=35370 / ATCC / 00234 / serotype MR
MEDLINE=21192684; PubMed=11296296;
RX

RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,

RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

RP SEQUENCE FROM N.A.

RX MEDLINE=22133808; Pubmed=12122206;

RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,

RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phage-encoded toxins, the high- virulence phenotype, and clone
 RT emergence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 DR EMBL: AE006576; AAK34209.1; -
 DR EMBL: AE014157; AAK79666.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C, PPIASE_2; 1.
 KW Hypothetical protein; Protease; Complete proteome.
 SO SEQUENCE 351 AA; 38533 MW; EE3A620A26EEF3A8 CRC64;

Query Match 43.8%; Score 696; DB 16; Length 351;
 Best Local Similarity 44.7%; Pred. No. 6,6e-30;
 Matches 140; Conservative 62; Mismatches 105; Indels 6; Gaps 2;

QY 13 KLLAGAITLLSVNLACSGSEGADLISKGVITEHOFYEQVKNPSAQVLLNMTQ 72
 Db 6 KLIASVYTLASVALAACQSTNDNTNIVISKGDTISVSDPYNETKTEVSQKMLNLV 65
 QY 73 KVEEKQYSGELDKEDVDITAEKKQYGENYQVLSOAGMTLETTRKAQIRTSKLVELAVK 132
 Db 66 RVEFAQYGDYKSKVEKAVHKTAEQYGFSAALAOSSLTPETFRKQIRSSKLVEYAVK 125
 QY 133 KVAEALTDEAYKAFDEYTPDYTAQIIRLNEDKAEVLEKAKAGADFAQLAKNDSTD 192
 Db 126 EAAKKELTQYKAYESYPTMAVEITLDNETAKSVLEELKAGADFTALAKEKTT- 184
 QY 193 EKTENGELTFPSASTVEPEQYKKAFAFDVDSVDTATGTAQVSSQYIVKLTCKT 252
 Db 185 --TPKKVYTKFDSGAINVPTDVYKKASSLNEGISDVISLDPSTYQKKFYIVKTKKA 242
 QY 253 EKSSNIDVYKREKTYILQKNDSTFVOSIIEKLEQANIKYKQAFONIFQYIGGD 312
 Db 243 EKSDQYKRRKATIIIAEKSDMNFQNKVIANALDKANVKKDKAFANITLQYANLQ 302
 QY 313 ---SSSSSTSN 322
 Db 303 KTKAASESTTSE 315

RESULT 7
 ID 08P0E5 PRELIMINARY; PRT; 351 AA.
 AC 08P0E5;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DE Putative protease maturation protein.
 GN spym18.1400.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 DR EMBL: AE010060; AAL97991.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C, PPIASE_2; 1.
 KW Protease; Complete proteome; Hypothetical protein.
 SO SEQUENCE 351 AA; 38533 MW; 5DB0499B1FA1BC9 CRC64;

Query Match 43.3%; Score 687; DB 16; Length 351;
 Best Local Similarity 44.18%; Pred. No. 2e-29;
 Matches 136; Conservative 63; Mismatches 106; Indels 6; Gaps 2;

QY 13 KLLAGAITLLSVNLACSGSEGADLISKGVITEHOFYEQVKNPSAQVLLNMTQ 72
 Db 6 KLIASVYTLASVALAACQSTNDNTNIVISKGDTISVSDPYNETKTEVSQKMLNLV 65
 QY 73 KVEEKQYSGELDKEDVDITAEKKQYGENYQVLSOAGMTLETTRKAQIRTSKLVELAVK 132
 Db 66 RVEFAQYGDYKSKVEKAVHKTAEQYGFSAALAOSSLTPETFRKQIRSSKLVEYAVK 125
 QY 133 KVAEALTDEAYKAFDEYTPDYTAQIIRLNEDKAEVLEKAKAGADFAQLAKNDSTD 192
 Db 126 EAAKKELTQYKAYESYPTMAVEITLDNETAKSVLEELKAGADFTALAKEKTT- 184
 QY 193 EKTENGELTFPSASTVEPEQYKKAFAFDVDSVDTATGTAQVSSQYIVKLTCKT 252
 Db 185 --TPKKVYTKFDSGAINVPTDVYKKASSLNEGISDVISLDPSTYQKKFYIVKTKKA 242
 QY 253 EKSSNIDVYKREKTYILQKNDSTFVOSIIEKLEQANIKYKQAFONIFQYIGGD 312
 Db 243 EKSDQYKRRKATIIIAEKSDMNFQNKVIANALDKANVKKDKAFANITLQYANLQ 302
 QY 313 ---SSSSSTSN 322
 Db 303 KTKAASESTTSE 315

RESULT 8
 ID 099XT9 PRELIMINARY; PRT; 309 AA.
 AC 099XT9;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Hypothetical protein SPY2037 (Hypothetical protein spym18.2097).
 GN SPY2037 OR spym18.2097.
 OS Streptococcus pyogenes, and
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314, 186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvarov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian X., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 DR EMBL: AE006625; AAK34705.1; -
 DR EMBL: AE010112; AAL98558.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C, PPIASE_2; 1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 309 AA; 34349 MW; 38DA6E0B27EALFE8 CRC64;

Query Match 39.0%; Score 619; DB 16; Length 309;

```
Best Local Similarity 42.6%; Pred. No. 7.5e-26;
Matches 127; Conservative 61; Mismatches 102; Indels 8; Gaps 3;

QY 13 KLLAGATITLVAATLACSGSGADLISMGKGVITEHQFYEQVKNPSAQOVLNMTIO 72
Db 6 KLTGGVTLATVATLACSSSHNNTKLVSMKGDITIVSDYNTKMTLQAOKMLSLVIS 65
QY 73 KYFEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETTRKAQIRTSKLVELAVK 132
Db 66 RYFEIOYANKVSDKEKAYKQPADYGTGTSFKTVLQASGLTPTTYKQIRTLTVLEYAVK 125
QY 133 KVAEAE-LTDEAVKKAFFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGADFAQLANDST 191
Db 126 EQAKNETISKDYRQAYDAYTPTMTAEIMQFEKEEDAKALEAVKAGADFAIAKEKTT 185
QY 192 --DEKTEKNGEITFPASASTEVEPOYKKAFAALDVGVSDVITATGTOAVSSQYTVKLT 249
Db 186 AADKKT-----TYTFSGETTLPAEYVRAASGLKEGNSREIITALDPATSKRTYHIKVT 240
QY 250 KTEKSSNIDYKEKELKTVILTKQKNDSTFVOSIIGKELQANIKYKQDAFQNIPTQY 307
Db 241 KATKKAADKAKAYOKRLKDIIVTGKLPDPFQNKVIKALDKANVKIKDKAFANILAQF 298

RESULT 9
Q8K5P3 PRELIMINARY; PRT; 309 AA.
AC 08K5P3:
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Putative protease maturation protein.
GN PRPM.2 OR SPYM3.1740.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=19846;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-MGAS315 / Serotype M3:
RA MEDLINE-22133808; PubMed-12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.:
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL: AE014170; AAM80347.1; -.
DR InterPro: IPR000297; Rotamase.
DR PROSITE: PS50198; PRIC_PRIASE_2; 1.
KW Protase; Complete protease; Hypothetical protein.
SQ SEQUENCE 309 AA; 34361 MW; 38DA7FEB2B1A02E8 CRC64;

Query Match 38.7%; Score 614; DB 16; Length 309;
Best Local Similarity 42.3%; Pred. No. 1.4e-25;
Matches 126; Conservative 61; Mismatches 103; Indels 8; Gaps 3;

QY 13 KLLAGATITLVAATLACSGSGADLISMGKGVITEHQFYEQVKNPSAQOVLNMTIO 72
Db 6 KLTGGVTLATVATLACSSSHNNTKLVSMKGDITIVSDYNTKMTLQAOKMLSLVIS 65
QY 73 KYFEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETTRKAQIRTSKLVELAVK 132
Db 66 RYFEIOYANKVSDKEKAYKQPADYGTGTSFKTVLQASGLTPTTYKQIRTLTVLEYAVK 125
QY 133 KVAEAE-LTDEAVKKAFFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGADFAQLANDST 189
Db 126 EQAKNETISKDYRQAYDAYTPTMTAEIMQFEKEEDAKALEAVKAGADFAIAKEKTT 185
QY 190 STDEKTEKNGEITFPASASTEVEPOYKKAFAALDVGVSDVITATGTOAVSSQYTVKLT 249
Db 186 AADKKT-----TYTFSGETTLPAEYVRAASGLKEGNSREIITALDPATSKRTYHIKVT 240

RESULT 10
Q8CEV9 PRELIMINARY; PRT; 308 AA.
AC 08CEV9:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Protein maturation protein.
GN PAPA OR IL1725.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_Taxid=1360;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-IL1403.
RA MEDLINE-21235186; PubMed-11337471;
RA Holcetin A., Wincker P., Manger S., Jalllon O., Malarme K.,
RA Weissendbach J., Ehrlich S.D., Sorokin A.:
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006403; AAK05823.1; -.
KW Complete proteome.
SQ SEQUENCE 308 AA; 33832 MW; F9E7FA832E9E0B00 CRC64;

Query Match 32.2%; Score 512; DB 16; Length 308;
Best Local Similarity 36.2%; Pred. No. 4e-20;
Matches 114; Conservative 65; Mismatches 120; Indels 16; Gaps 5;

QY 10 MKKKLAGATITL---SVATLACSGK--SEGADLISMGKGVITEHQFYEQVKNPS--A 62
Db 1 MKFKKLGIVMTYTFAGALVTLSCSSSDSASADITIMKGDITRVSDYKREKQFQSQT 60
QY 63 QOVLNMTIOKVFKEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETTRKAQIR 122
Db 61 NTLQNLITFDKIPKTFKFGKEVTDKDVSKKVSIKDQGSQFSSALQOQGLTEASFYPYMR 120
QY 123 TSKLVELAV-KKVAEDELDEAKKAFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGAD 181
Db 121 TOMLEQALPDHEIKETQYTDANLKKAMESYHPDVTAIVVSETSKDAATLADAKKDDAG 180
QY 182 FAQLANDNSTDEKTEKNGEITFPASASTEVEPOYKKAFAALDVGVSDVITATGTOAVSS 241
Db 181 KASFETKME-----SKVTFNSTSTSVPTFVQRAAFRLKNGEESDVIESSTGTANT 232
QY 242 QYTVKLTKEKTEKSSNIDYKEKELKTVILTKQKNDSTFVOSIIGKELQANIKYKQDAFQ 301
Db 233 SYIVEMVKTSEKGTDMNKRYKRELQNVIKTEKEQDTTFVSGVIAYKLKNNVYKESARA 292
QY 302 NITFOYIGGDDSSSS 316
Db 293 SLFSQFTQTSSSSSS 307

RESULT 11
Q929F4 PRELIMINARY; PRT; 291 AA.
AC 0929F4:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Hypothetical protein lin2322.
GN Listeria innocua.
OS Listeria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_Taxid=1642;
```

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CLIP 11262 / Serovar 6a;
 RA Glaser P., Francaul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Cherconul F., Couve E., de Darivar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Eutlian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,
 RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kunapat G.,
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueler T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of *Listeria species*."
 RL Science 294:849-852(2001).
 DR EMBL: AL596171; CAC97550.1; -
 DR Listlist: L1N02322; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIIC_PPIASE_2; 1.
 KM Hypothetical protein; Complete proteome.
 SO SEQUENCE 291 AA; 32359 MW; 85B0E3E5FC5A4FAL CRC64;

Query Match 25.28; Score 400.5; DB 16; Length 291;
 Best Local Similarity 33.68; Pred. No. 3.4e-14;
 Matches 103; Conservative 76; Mismatches 111; Indels 17; Gaps 11;

10 MKKKLAGATLLSVATLACSGSEBAGDLSMKGVITHEOFEOYKSNPSAQOYLNM 69
 1 MKKKLIGLVMMALFSLAAC--GGGNNVKTDSGDVDELXDAMKDKYGSFV-QQL 56
 70 TIOKVEKQYGESELDKREVDITAEKKQYGENYORVLSQAGMTLETRKAQIRTSKLV 129
 57 TFEKILGDKY--KVSDEDVKKFENEKSYQGDQFSALVSGLETKESFKSGLKYNLLVOK 114
 130 AVKVAEALITDEAYKKAPEYTPDYATQIIRLNEDKAEVLEKKAEGADPAQAKDN 189
 115 ATE--ANADTSDKALKEFYTPQPDITVSHILVADENKAEVDEKLA-DEAKPADLAKKY 171
 190 STDEKTEKNGEITFDSASTEPVEQYKKAFFALDVGVSDVITATGQVASSQYIYIKLT 249
 172 STDTATKENGQOLA-PFGSGKMDPAFEKAYALKKNG--DISAPVTKQ-YG--YHIQMD 225
 250 KTEKSSNIDYKELKTVILITOKNDSTFVQSIIGKELQANIKYKDAFQNIITQYIG 309
 226 KPATKTTFFDKRAKAEVLAQLTLEN--MORTLKEKVDANVKEKDKLDAFKDFDG 283
 310 GGDSSSS 316
 284 SASKSS 290

RESULT 12
 Q8Y557 PRELIMINARY; PRT; 293 AA.
 AC Q8Y557;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein lmo2219.
 GN LMO2219.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacilliales; Listeriaceae; *Listeria*.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Francaul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Darivar A., Dehoux P.,

Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Eutlian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,
 RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kunapat G.,
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueler T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of *Listeria species*."
 RL Science 294:849-852(2001).
 DR EMBL: AL591982; CAD00297.1; -
 DR Listlist: LMO02219; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIIC_PPIASE_2; 1.
 KM Hypothetical protein; Complete proteome.
 SO SEQUENCE 293 AA; 32698 MW; 172EDF44A8F7E7B CRC64;

Query Match 24.88; Score 393.5; DB 16; Length 293;
 Best Local Similarity 32.44; Pred. No. 8.2e-14;
 Matches 100; Conservative 77; Mismatches 115; Indels 17; Gaps 11;

10 MKKKLAGATLLSVATLACSGSEBAGDLSMKGVITHEOFEOYKSNPSAQOYLNM 69
 1 MKKKLIGLVMMALFSLAAC--GGGNNVKTDSGDVDELXDAMKDKYGSFV-QQL 56
 70 TIOKVEKQYGESELDKREVDITAEKKQYGENYORVLSQAGMTLETRKAQIRTSKLV 129
 57 TFEKILGDKY--KVSDEDVKKFENEKSYQGDQFSALVSGLETKESFKSGLKYNLLVOK 114
 130 AVKVAEALITDEAYKKAPEYTPDYATQIIRLNEDKAEVLEKKAEGADPAQAKDN 189
 115 ATE--ANADTSDKALKEFYTPQPDITVSHILVADENKAEVDEKLA-DEAKPADLAKKY 171
 190 STDEKTEKNGEITFDSASTEPVEQYKKAFFALDVGVSDVITATGQVASSQYIYIKLT 249
 172 STDTATKENGQOLA-PFGSGKMDPAFEKAYALKKNG--DISAPVTKQ-YG--YHIQMD 225
 250 KTEKSSNIDYKELKTVILITOKNDSTFVQSIIGKELQANIKYKDAFQNIITQYIG 309
 226 KPATKTTFFDKRAKAEVLAQLTLEN--MORTLKEKVDANVKEKDKLDAFKDFDG 283
 310 GGDSSSS 318
 284 SSSSDSS 292

RESULT 13
 Q9A103 PRELIMINARY; PRT; 299 AA.
 AC Q9A103;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Prtm precursor.
 GN Prtm.
 OS *Lactococcus lactis* (subsp. cremoris) (*Streptococcus cremoris*).
 OC Plasmid pHP003.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; *Lactococcus*.
 OX NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HP;
 RX MEDLINE=21367818; PubMed=11473599;
 RA Christensson C., Pillidge C.J., Ward L.J., O'Toole P.W.,
 RT "Nucleotide sequence and characterization of the cell envelope
 proteinase plasmid in *Lactococcus lactis* subsp. cremoris HP."
 RL J. Appl. Microbiol. 91:334-343(2001).
 DR EMBL: AF247159; AAK27980.1; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIIC_PPIASE_1; 1.
 DR PROSITE: PSS0198; PPIIC_PPIASE_2; 1.

FW	Signal.	POTENTIAL.
FT	1	
SEQUENCE	299 AA; 33211 MW; 20A160D4AD901F5B CRC64;	
Query Match	23.5%; Score 373; DB 2; Length 299;	
Best Local Similarity	31.7%; Pred. No. 1e-12;	
Matches	97; Conservative 59; Mismatches 128; Indels 22; Gaps 8;	
QY	10 MKK-----LLAGATLLSVATLAAACSGSGEADLISMGDVTTEHOFYEQVKSNPSSAQ 64	
DB	1 MKKKRLKVLVLASTATALL--LSQCSQNDQDQVATVSGKVTSESSFYKELKQSPTRKT 58	
QY	65 VLLNNTIOKVEKQYGESELDKREVDOTIAEKKQYGEYQYVLSQAGMTLETRKAQIRTS 124	
DB	59 MIANNLIYRALNHAYGKSVSTRVADADYSYKQOYGENFDALISONGFSRSSFKESIRTN 118	
QY	125 KLVELA---VKKVAEELTDEAYKKAPEYTPDVAQIIRLNNEPKAEVLEKKAEBAD 181	
DB	119 FLSEVALKKLKLVSSQL-----KAAVTYQPKYVQHILTSDEDTAKQVISDL-ASQKD 172	
QY	182 FAQLAKDSTDEKTEKGETTFDSASVEYEQYKKAFFALDVGDSVITATGTQAVSS 241	
DB	173 FAMLAKTSDIDTATFNDNGKTSIFELNNNTLTATFDDAYKTK---NDDTQTQPKV-TD 227	
QY	242 QYIYKLTTKTEKSSNIDYREKTKTYLLTQKNDSTFVSITIGELQAANIKVKQDAFQ 301	
DB	228 GYEYVKMINHPAKGT-FTSSKKVLIVASYAKWSRDSINQRYISQVLKNQHTIIXDKDLA 286	
QY	302 NIFTQY 307	
DB	287 DALDSY 292	
RESULT 14		
ID	053962	PRELIMINARY: PRT: 295 AA.
AC	053962:	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	
DR	01-NOV-1996 (TREMBLrel. 01, last sequence update)	
DE	01-JUN-2001 (TREMBLrel. 17, last annotation update)	
DE	Cremoris Wg2 proteinase (fragment).	
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.	
OX	NCBI_TaxId:1359;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-88149035: PubMed-3278687;	
RA	Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;	
RT	"Nucleotide sequence of the cell wall proteinase gene of Streptococcus cremoris Wg2.";	
RL	Appl. Environ. Microbiol. 54:231-238(1988).	
DR	EMBL: M24767; AAA17676.1; -	
DR	InterPro: IPR000297; Rotamase.	
DR	Pfam: PF00639; Rotamase; 1.	
DR	PROSITE: PS01096; PPIG_PPIASE_1; 1.	
DR	PROSITE: PSS0198; PPIG_PPIASE_2; 1.	
FT	NON_TER 295	
SO	SEQUENCE 295 AA; 32735 MW; A249053A1AB4FE7F CRC64;	
Query Match 22.9%; Score 363; DB 2; Length 295;		
Best Local Similarity 31.4%; Pred. No. 3.5e-12;		
Matches 96; Conservative 58; Mismatches 130; Indels 22; Gaps 8;		
QY	10 MKK-----LLAGATLLSVATLAAACSGSGEADLISMGDVTTEHOFYEQVKSNPSSAQ 64	
DB	1 MKKKRLKVLVLASTATALL--LSQCSQNDQDQVATVSGKVTSESSFYKELKQSPTRKT 58	
QY	65 VLLNNTIOKVEKQYGESELDKREVDOTIAEKKQYGEYQYVLSQAGMTLETRKAQIRTS 124	
DB	59 MIANNLIYRALNHAYGKSVSTRVADADYSYKQOYGENFDALISONGFSRSSFKESIRTN 118	
QY	125 KLVELA---VKKVAEELTDEAYKKAPEYTPDVAQIIRLNNEPKAEVLEKKAEBAD 181	
DB	119 FLSEVALKKLKLVSSQL-----KAAVTYQPKYVQHILTSDEDTAKQVISDL-ASQKD 172	
QY	182 FAQLAKDSTDEKTEKGETTFDSASVEYEQYKKAFFALDVGDSVITATGTQAVSS 241	
DB	173 FAMLAKTSDIDTATFNDNGKTSIFELNNNTLTATFDDAYKTK---NDDTQTQPKV-TD 227	
QY	242 QYIYKLTTKTEKSSNIDYREKTKTYLLTQKNDSTFVSITIGELQAANIKVKQDAFQ 301	
DB	228 GYEYVKMINHPAKGT-FTSSKKVLIVASYAKWSRDSINQRYISQVLKNQHTIIXDKDLA 286	
QY	302 NIFTQY 307	
DB	287 DALDSY 292	
RESULT 14		
ID	053962	PRELIMINARY: PRT: 295 AA.
AC	053962:	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	
DR	01-NOV-1996 (TREMBLrel. 01, last sequence update)	
DE	01-JUN-2001 (TREMBLrel. 17, last annotation update)	
DE	Cremoris Wg2 proteinase (fragment).	
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.	
OX	NCBI_TaxId:1359;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-88149035: PubMed-3278687;	
RA	Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;	
RT	"Nucleotide sequence of the cell wall proteinase gene of Streptococcus cremoris Wg2.";	
RL	Appl. Environ. Microbiol. 54:231-238(1988).	
DR	EMBL: M24767; AAA17676.1; -	
DR	InterPro: IPR000297; Rotamase.	
DR	Pfam: PF00639; Rotamase; 1.	
DR	PROSITE: PS01096; PPIG_PPIASE_1; 1.	
DR	PROSITE: PSS0198; PPIG_PPIASE_2; 1.	
FT	NON_TER 295	
SO	SEQUENCE 295 AA; 32735 MW; A249053A1AB4FE7F CRC64;	
Query Match 22.9%; Score 363; DB 2; Length 295;		
Best Local Similarity 31.4%; Pred. No. 3.5e-12;		
Matches 96; Conservative 58; Mismatches 130; Indels 22; Gaps 8;		
QY	10 MKK-----LLAGATLLSVATLAAACSGSGEADLISMGDVTTEHOFYEQVKSNPSSAQ 64	
DB	1 MKKKRLKVLVLASTATALL--LSQCSQNDQDQVATVSGKVTSESSFYKELKQSPTRKT 58	
QY	65 VLLNNTIOKVEKQYGESELDKREVDOTIAEKKQYGEYQYVLSQAGMTLETRKAQIRTS 124	
DB	59 MIANNLIYRALNHAYGKSVSTRVADADYSYKQOYGENFDALISONGFSRSSFKESIRTN 118	
QY	125 KLVELA---VKKVAEELTDEAYKKAPEYTPDVAQIIRLNNEPKAEVLEKKAEBAD 181	
DB	119 FLSEVALKKLKLVSSQL-----KAAVTYQPKYVQHILTSDEDTAKQVISDL-ASQKD 172	

Dd		119	FLSEVALKKLKVSESOL----	KAAKTYPPKYTVQHILTSDEDTAKQVIVSLD-AAKG	172
Oy		182	FAGLAQNDSDEKTRENGEGETTFPDSASTVEPYVKKAFAFALVDGVSDVIATGTQAYS	241	
Dd		173	FAMIAKTSDIDTAIRKNCGKISFPENLNNTLTDAFPAKDAAYKLK-----NODYIQTFPVK--TD	227	
Oy		242	OYIYVLTKTRKESSINIDDYEKELVTYLITOKUNDSTVFVSIIIGKELOAANIKRKDOAFQ	301	
Dd		228	GVEYIKMHPBPKGT-FITSRKKALTASVYAARWMSDBSJMOPVISOVUKNKOHVTTIKDKDLA	286	
Oy		302	NIFTOY 307		
Dd		287	DALDY 292		
		RESULT 15			
ID	O8Y759	PRELIMITARY:	PRT;	294 AA.	
AC	O8Y759				
DR	01-MAR-2002 (TREMBLrel. 20,	Created)			
DR	01-MAR-2002 (TREMBLrel. 20,	Last sequence update)			
DI	01-JUN-2002 (TREMBlrel. 21,	Last annotation update)			
DE	Hypothetical protein Imo1444.				
CN	LMO1444.				
OS	Listeria monocytogenes.				
OC	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.				
OX	NCBI_Taxid=1639;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EGB-e / Serovar 1/2a;				
RX	MEDLINE=21537279; PubMed=11679669;				
RA	Glaser P., Frangeul L., Buchrieser C., Rusnlok C., Amend A.,				
RA	Baquero F., Berche P., Blocher H., Brandt P., Chakraborty T.,				
RA	Charbit A., Checouani F., Couve E., de Daruvay A., Delhoux P.,				
RA	Domann E., Dominguez-Bernal G., Duchud E., Durant L., Hussirget O.,				
RA	Eutlian K.-D., Fstini W., Garcia-del Portillo F., Garrido P.,				
RA	Gautier I., Goebel M., Gomez-Lopez N., Halin T., Hauf J., Jackson D.,				
RA	Jones L.-M., Kaestli U., Kreft J., Kunz M., Kunst F., Kurapkat G.,				
RA	Madueno E., Maltonnam A., Mata Vicente J., Ng E., Nedjari H.,				
RA	Nordstad G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,				
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,				
RA	Vazquez-Boland J.-A., Voss H., Weiland J., Cosart P.;				
RT	"Comparative genomics of Listeria species";				
RL	Science 294:849-852(2001).				
DR	EMBL; AL591979; CAC99522.1; -				
DR	Listlist: IMO01444; -				
DR	InterPro: IPRO00297; Rotamase.				
DR	Pfam; PF00639; Rotamase; 1.				
DR	PROSITE; PS50198; PRIC_PPASE_2; 1.				
SQ	Hypothetical proteome. Complete proteome.				
KW	SEQUENCE 294 AA; 32569 MW; 190C705AUBFJIEB7 CRC64;				
	Query Match	20.7%;	Score 328;	DB 16; Length 294;	
	Best Local Similarity	30.7%;	Pred. No. 2.6e+10;		
	Matches 96; Conservative	71; Mismatches 120;	Indels 26; Gaps 14;		
Oy	12	KKLAGAT--TLTSVAATIACSCKSGEGADLSMGKDVTTEHOFYOVKSNSPSAQOVLNM	69		
Dd	5	KKVMISVIAAILL---LLAGCSSA----VKTGASTYTODELYIAMATT-YGNFYVOL	56		
Oy	70	TIOKFVFEKOYGSELDDKEVEDPTIAEEKROYGENVOAVLSQAGMTLEFRKAOIRTSKVEL	129		
Dd	57	TFPKLIKLEDK--TVTEKEVNAEYKKYEBOGDSPFTLISSNNLKTIETSKPENLEYLVLLVOK	114		
Oy	130	A VKKAAEEMLIDEAYKKAFFDETTPVTYKQIIRLNNEDEKAREYLEKAKEGADFOLAOND	189		
Dd	115	ATE--ANNADVSESKLAYKKTWEPDITHRIHALVIDEAETAKEIoTQLK-NGERFTDLAREY	171		
Oy	190	STDFKTENGGETTIPDSASTEPEDEVOKKAAFAFD-VPGSVPIATATQTQAVSSOYYIVKL	248		
Dd	172	STDYATSTING-LLDPFGEKHDEFERKAAHYALEKKDVSGIVKST-----YG-YHLQOL	224		
Oy	249	TKTKRSSNIIDYKERELKTVIJLITORUNDSTFVASIIGKELOAANIKRVKDAQFOINIITYOI	308		

Fri Sep 5 09:24:51 2003

us-10-049-473a-2.rspt

Page 8

Db	225	VKTEGTAKE-KANVKAAT-KSOLISENNATKREIKKANIDIKSDLKDAFYD	282
Qy	309	GGDSSSSSTSN	321
Db	283	-STSSTSTTSN	294

Search completed: August 29, 2003, 10:36:13
Job time : 104 secs

Fri Sep 5 09:24:49 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:11:17; Search time 3649 Seconds

(without alignments)
10630.008 Million cell updates/sec

Title: US-10-049-473A-1

Perfect score: 966

Sequence: 1 agtaacactatctcaag.....gcagtagatcacaacgaa 966

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenBankl:*

1: gb_ba:*

2: gb_btq:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htg_hum:*

40: em_htg_mus:*

41: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	966	6	AX085885 Sequence
2	966	100.0	966	6	AX138385 Sequence
3	966	100.0	11309	6	AR218876 Sequence
4	966	100.0	11309	6	BD003788 Polynucle
5	966	100.0	13378	1	AE007401 Streptoco
6	966	100.0	34980	6	AX571762 Sequence
7	961.2	99.5	11552	1	AE008462 Sequence
8	939	97.2	939	6	AX568726 Sequence
9	939	97.2	111135	2	SPNEU1906
10	552.4	57.2	640	6	AR120254 Sequence
11	552.4	57.2	640	6	BD063263 Streptoco
12	265.6	27.5	791	6	AR193967 Sequence
13	265.4	27.5	12012	1	AE014808 Streptoco
14	261.6	27.1	11615	1	AE006576 Streptoco
15	260	26.9	51454	1	AE014157 Streptoco
16	259.4	26.9	310950	1	AP005143 Streptoco
17	259.4	26.9	20639	1	AE014228 Streptoco
18	257.8	26.7	930	6	AX607662 Sequence
19	257.8	26.7	6641	6	AX602145 Sequence
20	257.8	26.7	167050	1	SA6766847 Streptoco
21	255.2	26.4	10844	1	AE010060 Streptoco
22	216.2	22.4	1110	6	AR228534 Sequence
23	215.6	22.3	300225	1	AE016949 Streptoco
24	211.6	21.9	10029	1	AE010112 Streptoco
25	211.6	21.9	12808	1	AE006625 Streptoco
26	210	21.7	33291	1	AE014170 Streptoco
27	210	21.7	323825	1	AP005146 Streptoco
28	203.2	21.0	970	6	AR228535 Sequence
29	150.4	15.6	12898	1	AE006403 Streptoco
30	125	12.9	201688	2	AC142194 Homo sapi
31	91.2	9.4	302050	1	AL935256 Lactobaci
32	84	8.7	232807	2	SPNEU1901 Streptoco
33	76.6	7.9	43980	6	AX413014 Sequence
34	76.6	7.9	258650	1	AL596171 Listeria
35	76.6	7.9	349980	6	AX417047 Sequence
36	73.8	7.6	3720	14	HSV3BGEN
37	73.8	7.6	43658	14	HSV3BGEN
38	73.8	7.6	112830	14	HSV3BGEN
39	72	7.5	295050	1	AL591982 Listeria
40	72	7.5	34980	6	AX641671 Sequence
41	71.8	7.4	825	3	AF206632 Plasmodiu
42	71.8	7.4	1720	6	AX416363 Sequence
43	70.8	7.3	110000	3	PRMAL1P2_0 ContInnatio
44	70.2	7.3	110000	2	PRMAL13_07 ContInnatio
45	67.8	7.0	304708	1	AE017001 Bacillus

ALIGNMENTS

RESULT 1	AX085885	966 bp	DNA	linear	PAT 09-MAR-2001
LOCUS	AX085885				
DEFINITION	Sequence 1 from Patent WO0112219.				
ACCESSION	AX085885				
VERSION	AX085885.1	GI:13275796			
KEYWORDS					
SOURCE	Streptococcus pneumoniae				
ORGANISM	Streptococcus pneumoniae				
REFERENCE	1				
AUTHORS	de Groot, R. and Hermans, P.W.				
TITLE	Pneumococcal vaccines				
JOURNAL	Patent: WO 0112219-A 1 22-FEB-2001;				

FEATURES	Erasmus	Location/Qualifiers
SOURCE	1. .966	
	/organism="Streptococcus pneumoniae"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:1313"	
	28. .966	
CDS	/note="unnamed protein product"	
	/codon_start=1	
	/transl_table=1	
	/protein_id="CAC34049.1"	
	/db_xref="gi:13275797"	
	/translation="MKKKLAGATTLISVATLACSKSGEGADLISMKGVIPTHEPXY	
	EOVSNPSAOOVLTLNMTIOKVEFKOYSESLDEKVDPTIAEKKQYSENRQRTISQAG	
	MTLETRRAQITSLKLYELAVKAYVAEALTEIDAKRDEITPDVTAQIIRLNNDKAK	
	EVLKAKAEGDAFQALKDNDSTDEKTEKNGEITFDSASTPEPQVAKAAFLVDVQ	
	SDVATIGQTAIVSSQYIVITVILTKTKTERSSNIDVKEKLYVILLQKNDSTFVSIQIG	
	KETLDAAIRKVDQAFQNIFFQYIGGGSSSSSSSTNSNE"	
BASE COUNT	374 a	170 c 187 g 235 t
ORIGIN		

Query Match	Similarity	100.0%	Score 966:	DB 6:	Length 966:
Best Local	Similarity	100.0%	Pred. NO. 2.1e-167:		
Matches	966:	Conservative	0:	Mismatches	0:
				Indels	0:
				Gaps	0:
QY	1	AGTAACTATATCCAAAGAGTAGACATGAAAGAAAAATATTATGGCAGGTGCCATACA	60		
Db	1	AGTAACTATATCCAAAGAGTAGACATGAAAGAAAAATATTATGGCAGGTGCCATACA	60		
QY	61	CTATTATCAGTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAAAGTGCAGACTTATC	120		
Db	61	CTATTATCAGTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAAAGTGCAGACTTATC	120		
QY	121	AGCATGAAAGGGGATGTCATTCAGAACATCAATTTATATGCGAAGTGAAGAAAGAACCT	180		
Db	121	AGCATGAAAGGGGATGTCATTCAGAACATCAATTTATATGCGAAGTGAAGAAAGAACCT	180		
QY	181	TCAGCCCAACAGCTCTGTTTAAATATGACCATCCAAAGATTTTGAAGAAACATATGGC	240		
Db	181	TCAGCCCAACAGCTCTGTTTAAATATGACCATCCAAAGATTTTGAAGAAACATATGGC	240		
QY	241	TCAGAGCTTGATGATTAAGAGGTTGATGATCTTTCGCGAAGAAAAACAATATGGC	300		
Db	241	TCAGAGCTTGATGATTAAGAGGTTGATGATCTTTCGCGAAGAAAAACAATATATGGC	300		
QY	301	GAAACTACCAACGCTCTCTTCACAAGCAGTAGTGACTCTGAAACACGTAAAGCTCAA	360		
Db	301	GAAACTACCAACGCTCTCTCTTCACAAGCAGTAGTGACTCTGAAACACGTAAAGCTCAA	360		
QY	361	ATTTCGTAAGTAATTAATAGTTGAGTTGGCAGATTAAAGAGTACGAGCAAGCTGAATTGACA	420		
Db	361	ATTTCGTAAGTAATTAATAGTTGAGTTGGCAGATTAAAGAGTACGAGCAAGCTGAATTGACA	420		
QY	421	GATCAACGCTTATAGAAAGCCTTTGATGATGATCACTCCAGATGTATACGGCTCAATATATC	480		
Db	421	GATCAACGCTTATAGAAAGCCTTTGATGATGATCACTCCAGATGTATACGGCTCAATATATC	480		
QY	481	CGTCTTATATGTAAGATTAAGGCCCAAGAGTTTCGAAAAAGCCCAAGGCAAGAGTGC	540		
Db	481	CGTCTTATATGTAAGATTAAGGCCCAAGAGTTTCGAAAAAGCCCAAGGCAAGAGTGC	540		
QY	541	GATTTTCTCATTTAGCCAAAGATTAATTCACAGTGAATAAACAAGAAATATGATGGA	600		
Db	541	GATTTTCTCATTTAGCCAAAGATTAATTCACAGTGAATAAACAAGAAATATGATGGA	600		
QY	601	GAAATTCAGTTGATTTGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC	660		
Db	601	GAAATTCAGTTGATTTGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC	660		
QY	661	GCTTAAATGAGATGAGTGTTCGATGATGATTAACAGCACTGGCACAACAAGCTACAGT	720		
Db	661	GCTTAAATGAGATGAGTGTTCGATGATGATTAACAGCACTGGCACAACAAGCTACAGT	720		

QY	721	AGCCAAATTTACATGTGTAAGACTCACTAGAGAAACAGAAATATCAATTAATTTGATGAC	780
Db	721	AGCCAAATTTACATGTGTAAGACTCACTAGAGAAACAGAAATATCAATTAATTTGATGAC	780
QY	781	TACAAAGAAAAATTTAAAAAAGTGTATCTTGACTCAAAAAACAAATGATTCACATTTGTT	840
Db	781	TACAAAGAAAAATTTAAAAAAGTGTATCTTGACTCAAAAAACAAATGATTCACATTTGTT	840
QY	841	CAAGCATTATCGGAAAGAAATTTGCAAGCAGCCATTATCAAGTTAAGSACCAAGCCCTTC	900
Db	841	CAAGCATTATCGGAAAGAAATTTGCAAGCAGCCATTATCAAGTTAAGSACCAAGCCCTTC	900
QY	901	CAAAATATCTTTCACCAATATATTCGGTGGTGAGATTCAAGCTCAAGCAGTAGTACATCA	960
Db	901	CAAAATATCTTTCACCAATATATTCGGTGGTGAGATTCAAGCTCAAGCAGTAGTACATCA	960
QY	961	AACGAA 966	
Db	961	AACGAA 966	

LOCUS	AX138385	966 bp	DNA	linear	PAT 30-MAY-2001
DEFINITION	Sequence 1 from Patent EP1075841.				
ACCESSION	AX138385				
VERSION	AX138385.1	GI:14274335			
KEYWORDS					
SOURCE					
ORGANISM	Streptococcus pneumoniae Streptococcus pneumoniae Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1				
AUTHORS	de Groot, R. and Hermans, P. W.				
TITLE	Pneumococcal vaccines				
JOURNAL	Patent: EP 1075841-A 1 14-FEB-2001; Erasmus Universiteit Rotterdam (NL)				
FEATURES	Location/Qualifiers				
source	1..966 /organism="Streptococcus pneumoniae" /mol_type="genomic DNA" /db_xref="taxon:1313" /note="The S. pneumoniae nucleotides 7632-8597 on contig 33"				
BASE COUNT	374 a 170 c 187 g 235 t				
ORIGIN					
Query Match	100.0%; Score 966; DB 6; Length 966;				
Best Local Similarity	100.0%; Pred. No. 2.1e-167;				
Matches	966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 AGTAACACTTTCGCAAGAGAGTAGACATGAAAGAAAATTTATTCGACAGTCCATCA 60				
Db	1 AGTAACACTTTCGCAAGAGAGTAGACATGAAAGAAAATTTATTCGACAGTCCATCA 60				
OY	61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGAGTGCAGACCTTATC 120				
Db	61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGAGTGCAGACCTTATC 120				
OY	121 AGCATGAAAGGGGATGTCATTACGACACATCAATTTTATGAGCAAGTAAAGCAACCT 180				
Db	121 AGCATGAAAGGGGATGTCATTACGACACATCAATTTTATGAGCAAGTAAAGCAACCT 180				
OY	181 TCAGCCCAACAGCTCTGTTAAATATGACATCCCAAAAAGTTTTGAAAAACAATATGGC 240				
Db	181 TCAGCCCAACAGCTCTGTTAAATATGACATCCCAAAAAGTTTTGAAAAACAATATGGC 240				
OY	241 TCAGAGCTTGATGATAAGAGGTGATATCTATTTGCCAAGAAAAACAATATGGC 300				
Db	241 TCAGAGCTTGATGATAAGAGGTGATATCTATTTGCCAAGAAAAACAATATGGC 300				
OY	301 GAAATATACCAAGCTGTCTTGTCCACAGCAGATGATCTTTGAAACAGTAAAGCTCA 360				
Db	301 GAAATATACCAAGCTGTCTTGTCCACAGCAGATGATCTTTGAAACAGTAAAGCTCA 360				

DB 301 GAAACACCAAGCTGCTTGTCTACAGCAGGTATGACTCTTGAAACAGCTAAAGCTCA 360
QY 361 ATTCGTACAGTAATTAAGTTAGTGGCAGTTAAGAGTACAGAAAGCTGAATTGACA 420
DB 361 ATTCGTACAGTAATTAAGTTAGTGGCAGTTAAGAGTACAGAAAGCTGAATTGACA 420
QY 421 GATGAGCCTATAGAAAGCCTTGTATGATGACTCCAGATGTAAAGGCTCAAAATCATC 480
DB 421 GATGAGCCTATAGAAAGCCTTGTATGATGACTCCAGATGTAAAGGCTCAAAATCATC 480
QY 481 CGCTTAATTAAGATTAAGGCCAAAGGTTCTCGAAAGGCAAGGCCAAGCAAGTGT 540
DB 481 CGCTTAATTAAGATTAAGGCCAAAGGTTCTCGAAAGGCAAGGCCAAGCAAGTGT 540
QY 541 GATTTGCTCAATTAAGCCAAAGATATCACTGATGAGAAAGCAAGAAAGTGTGTA 600
DB 541 GATTTGCTCAATTAAGCCAAAGATATCACTGATGAGAAAGCAAGAAAGTGTGTA 600
QY 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC 660
DB 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC 660
QY 661 GCTTAGATGAGTGGAGTGTCTTCTGATGATGATGATGATGATGATGATGATGAT 720
DB 661 GCTTAGATGAGTGGAGTGTCTTCTGATGATGATGATGATGATGATGATGATGAT 720
QY 721 AGCCAAATATTAATTAAGTAACTCACTAAGAAAGCAAGAAATCACTAATTAATGATGAC 780
DB 721 AGCCAAATATTAATTAAGTAACTCACTAAGAAAGCAAGAAATCACTAATTAATGATGAC 780
QY 781 TACAAAGAAAAATTAAGAACTGTATCTTGACTCAAAAACAAAATGATTAATGATGTT 840
DB 781 TACAAAGAAAAATTAAGAACTGTATCTTGACTCAAAAACAAAATGATTAATGATGTT 840
QY 841 CAACACATTAATGAGAAAGATTAAGCAAGCAATATCAAGTTAAGGCAAGCAAGCTTC 900
DB 841 CAACACATTAATGAGAAAGATTAAGCAAGCAATATCAAGTTAAGGCAAGCAAGCTTC 900
QY 901 CAAAATATCTTTACCAATATATCGTGTGAGATTCAGCTCAAGCTAGTATCATCA 960
DB 901 CAAAATATCTTTACCAATATATCGTGTGAGATTCAGCTCAAGCTAGTATCATCA 960
QY 961 AACGAA 966
DB 961 AACGAA 966

RESULT 3
AR218876 11309 bp DNA linear PAT 25-SEP-2002
LOCUS AR218876
DEFINITION Sequence 108 from patent US 6420135.
ACCESSION AR218876
VERSION AR218876.1 GI:23319810
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 11309)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.S., Rosen,C.A., Barash,S.C.,
Fannon,M.R. and Dougherty,B.A.
TITLE Streptococcus pneumoniae polynucleotides and sequences
JOURNAL Patent: US 6420135-A 108 16-JUL-2002;
FEATURES
source 1. 11309
BASE COUNT 3557 a 2073 c 2451 g 3228 t
ORIGIN

Query Match 100.0%; Score 966; DB 6; Length 11309;
Best Local Similarity 100.0%; Pred. No. 1,2e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACACTTATCTCAAGAGTAGACATGAAGAAAAATTAATGCGAGTGCCATCACA 60

DB 7632 AGTAACACTTATCTCAAGAGTAGACATGAAGAAAAATTAATGCGAGTGCCATCACA 7691
QY 61 CTAATATACGTAGCAACTTATACAGCTTGTGAAAGGTCGAGAGTGACAGCTTATC 120
DB 7692 CTAATATACGTAGCAACTTATACAGCTTGTGAAAGGTCGAGAGTGACAGCTTATC 7751
QY 121 AGCATGAAAGGGGATGCTATTACAGAACATCAATTTTATGACAAAGTGAAGAACCT 180
DB 7752 AGCATGAAAGGGGATGCTATTACAGAACATCAATTTTATGACAAAGTGAAGAACCT 7811
QY 181 TCAGCCCAACAAAGTCTTGTAAATATGACATCCAAAAAGTTTGAAGAAATATGCG 240
DB 7812 TCAGCCCAACAAAGTCTTGTAAATATGACATCCAAAAAGTTTGAAGAAATATGCG 7871
QY 241 TCAGAGCTTGATGATTAAGAGGTGTGATGATGATGATGATGATGATGATGATGAT 300
DB 7872 TCAGAGCTTGATGATTAAGAGGTGTGATGATGATGATGATGATGATGATGATGAT 7931
QY 301 GAAACACTACCAAGCTGTCTTGTACAGAGCAGTATGACTCTTGAACACAGTAAAGCTCA 360
DB 7932 GAAACACTACCAAGCTGTCTTGTACAGAGCAGTATGACTCTTGAACACAGTAAAGCTCA 7991
QY 361 ATTCGTACAGTAATTAAGTTAGTGGCAGTTAAGAGGTACGAGAGCTGAATTGACA 420
DB 7992 ATTCGTACAGTAATTAAGTTAGTGGCAGTTAAGAGGTACGAGAGCTGAATTGACA 8051
QY 421 GATGAGCCTATTAAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGAT 480
DB 8052 GATGAGCCTATTAAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGAT 8111
QY 481 CGCTTAATTAATGAAGATTAAGGCCAAAGAGTCTCGAAAAAGCCAAAGCAGAGTGT 540
DB 8112 CGCTTAATTAATGAAGATTAAGGCCAAAGAGTCTCGAAAAAGCCAAAGCAGAGTGT 8171
QY 541 GATTTGCTCAATTAAGCCAAAGATTAATTAAGTGAAGAAAAACAAAGAAATGATGTA 600
DB 8172 GATTTGCTCAATTAAGCCAAAGATTAATTAAGTGAAGAAAAACAAAGAAATGATGTA 8231
QY 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC 660
DB 8232 GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC 8291
QY 661 GCTTAGATGAGTGGAGTGTCTTCTGATGATGATGATGATGATGATGATGATGAT 720
DB 8292 GCTTAGATGAGTGGAGTGTCTTCTGATGATGATGATGATGATGATGATGATGAT 8351
QY 721 AGCCAAATATTAATTAAGTAACTCACTAAGAAACAGAAATTCATCTAATTTGATGAC 780
DB 8352 AGCCAAATATTAATTAAGTAACTCACTAAGAAACAGAAATTCATCTAATTTGATGAC 8411
QY 781 TACAAAGAAAAATTAAGAACTGTATCTTGACTCAAAAACAAAATGATTAATGATGTT 840
DB 8412 TACAAAGAAAAATTAAGAACTGTATCTTGACTCAAAAACAAAATGATTAATGATGTT 8471
QY 841 CAACGATTAATGAGAAAGATTAAGTCAAGCAGCAATATCAAGTTAAGGCAAGCCTTC 900
DB 8472 CAACGATTAATGAGAAAGATTAAGTCAAGCAGCAATATCAAGTTAAGGCAAGCCTTC 8531
QY 901 CAAAATATCTTTACCAATATATCGTGTGAGATTCAGCTCAAGCTAGTATCATCA 960
DB 8532 CAAAATATCTTTACCAATATATCGTGTGAGATTCAGCTCAAGCTAGTATCATCA 8591
QY 961 AACGAA 966
DB 8592 AACGAA 8597

RESULT 4
BD003788 11309 bp DNA linear PAT 31-JAN-2002
LOCUS BD003788
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003788

```

VERSION      BD003788.1  GI:18631749
KEYWORDS     JP 2001501833-A/108.
SOURCE       unidentified
ORGANISM     Streptococcus pneumoniae
REFERENCE    1 (bases 1 to 11309)
AUTHORS      Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
              Fannon, M., and Dougherty, B.A.
TITLE        Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL      Patent: JP 2001501833-A 108 13-FEB-2001;
              HUMAN GENOME SCIENCES INC
COMMENT      OS Unidentified
              PN JP 2001501833-A/108
              PD 13-FEB-2001
              PF 30-OCT-1997 JP 1998520718
              PR 31-OCT-1996 US 60/029960
              PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
              STEVEN C BARASH,
              PI MICHAEL FANNON, BRIAN A DOUGHERTY
              PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
              PC C12N1/21,
              PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
              CC Strandedness: Double;
              CC Topology: Linear;
              FH Key
              FT source
              FT Location/Qualifiers
              1..11309
              /organism="unidentified"
              /mol_type="genomic DNA"
              /db_xref="taxon:32644"

BASE COUNT   3557 a 2073 c 2451 g 3228 t
ORIGIN
Query Match      100.0% Score 966; DB 6; Length 11309;
Best Local Similarity 100.0%; Pred. No. 1.2e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACACTTATCTCAAGAGTAGACATGAGAGAAAAATTATATGCGAGTCCATACACA 60
DB 7632 AGTAACTTATCTCAAGAGTAGACATGAGAGAAAAATTATATGCGAGTCCATACACA 7691
QY 61 CTATTATAGTAGACACTTTAGCAGCTTGTTCGAAAAGGTCGAGAGGTGCAGACCTTATC 120
DB 7692 CTATTATAGTAGACACTTTAGCAGCTTGTTCGAAAAGGTCGAGAGGTGCAGACCTTATC 7751
QY 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGAGCAGTGAAGAACCAACCT 180
DB 7752 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGAGCAGTGAAGAACCAACCT 7811
QY 181 TCAGCCCAACAAGTCTTGTAAATATGACATCCAAAAAGTTTGTGAAGAACCAATATGAC 240
DB 7812 TCAGCCCAACAAGTCTTGTAAATATGACATCCAAAAAGTTTGTGAAGAACCAATATGAC 7871
QY 241 TCAGAGCTTGATGATTAAGAGGTTGATGATCTATGCGGAGAAAGAAAAACAAATATGAC 300
DB 7872 TCAGAGCTTGATGATTAAGAGGTTGATGATCTATGCGGAGAAAGAAAAACAAATATGAC 7931
QY 301 GAAACATCAACAGCGTCTTGACACAGAGGATGACCTTGAAGAACGTTAAAGCTCAA 360
DB 7932 GAAACATCAACAGCGTCTTGACACAGAGGATGACCTTGAAGAACGTTAAAGCTCAA 7991
QY 361 ATTCGTACAGTAATTAATAGTTGAGTTGCGAGTTAAGAGGATGACGAGGCTGAATGACA 420
DB 7992 ATTCGTACAGTAATTAATAGTTGAGTTGCGAGTTAAGAGGATGACGAGGCTGAATGACA 8051
QY 421 GATGAAGCTTATAGAAAGCCTTTATGATGATGACTCCAGATGTAAGCGCTCAATATC 480
DB 8052 GATGAAGCTTATAGAAAGCCTTTATGATGATGACTCCAGATGTAAGCGCTCAATATC 8111
QY 481 CGTCTTAATTAATGAAGATTAAGGCCAAAGAAGTTCTGAAAAAGCCAAAGCAGACAGTCT 540

```

```

DB 8112 CGTCTTAATTAATGAAGATTAAGGCCAAAGAAGTTCTGAAAAAGCCAAAGCAGACAGTCT 8171
QY 541 GATTTGCTCATTTAGCCCAAGATTAATCAATGATGAGAAAAACAAAGAAAAATGCTGA 600
DB 8172 GATTTGCTCATTTAGCCCAAGATTAATCAATGATGAGAAAAACAAAGAAAAATGCTGA 8231
QY 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCCTTTC 660
DB 8232 GAAATTAACCTTGAATTCGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCCTTTC 8291
QY 661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 8292 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8351
QY 721 AGCCATATTTACATTTGATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 780
DB 8352 AGCCATATTTACATTTGATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 8411
QY 781 TACAAAGAAAAATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 840
DB 8412 TACAAAGAAAAATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 8471
QY 841 CAAGCATTTATCGGAAAAAGATTCGACAGCAGCATTTATCAAGTTAAGGACCAACCTTC 900
DB 8472 CAAGCATTTATCGGAAAAAGATTCGACAGCAGCATTTATCAAGTTAAGGACCAACCTTC 8531
QY 901 CAAATATCTTTACCAATATATCGGTGAGATTCGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 960
DB 8532 CAAATATCTTTACCAATATATCGGTGAGATTCGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 8591
QY 961 AACGAA 966
DB 8592 AACGAA 8597

RESULT 5
LOCUS      AE007401
DEFINITION Streptococcus pneumoniae TIGR4 section 84 of 194 of the complete genome.
ACCESSION  AE007401 AE005672
KEYWORDS   Streptococcus pneumoniae TIGR4
SOURCE     Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
ORGANISM   Streptococcus pneumoniae
REFERENCE  1 (bases 1 to 13378)
AUTHORS    Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
              Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
              Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
              Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
              Holtzapple, E., Khouri, H., Wolf, A.M., Ullrich, T.R., Hansen, C.L.,
              McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,
              Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,
              Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and
              Fraser, C.M.
TITLE      Complete genome sequence of a virulent isolate of Streptococcus pneumoniae
JOURNAL    Science 293 (5529), 498-506 (2001)
MEDLINE    21357209
PUBMED     11463916
REFERENCE  2 (bases 1 to 13378)
AUTHORS    Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
              Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
              Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
              Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E.,
              Khouri, H., Wolf, A.M., Ullrich, T.R., Hansen, C.L., McDonald, L.A.,
              Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E.,
              Loftus, B.J., Uwal, M.L., Yang, F., Smith, H.O., Venter, J.C.,
              Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
TITLE      Direct Submission

```

JOURNAL Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1. 13378

/organism="Streptococcus pneumoniae TIGR4"

/mol_type="genomic DNA"

/strain="TIGR4"

/db_xref="taxon:170187"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

83. .580

/gene="SP0967"

/note="identified by match to PFM protein family HM

PF01121"

/codon_start=1

/transl_table=1

/product="kinase, putative"

/protein_id="AAK75092.1"

/db_xref="GI:14972446"

/translation="MGKIGITGGIAGSKSTVNFPLRQGGFQVADAVYHQLQPG

RLFAVNOHFGQETIENGELNPLASLIFSNDPERMSKIOGELIREPLAREQ

LAOTEIIFPMIPLFEDQYSDWFAETLYVYDDAOVERLAKKRDQSKDEBSRLAA

QWPLEKKRDLSQVLDNNGNONLNOVHILLEGGRDDRD"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

/gene="SP0972"

3337. .4536

```

gene
CDS
    7377..7844
    /gene="Sp0976"
    /note="S0976"
    /note="identified by match to TIGR protein family HMM
    TIGR0086"
    /codon_start=1
    /transl_table=11
    /product="SsrA-binding protein"
    /protein_id="AAK75097.1"
    /db_xref="gi:14972451"
    /translation="MAKGEKRVVAONKRRARDYITVDLLEAGMVLITGEIKSVRAARI
    NLKDFACVKNKEVWLSNVHIAIPEEGNIMNOPEERKRLIHKROIOKLOEIKGTG
    MTLVPLKYIKDGYAKLLGLAKGHVDKNESTRRRONDIAIVKMAVNR"
    7860..8720

Query Match      100.0%; Score 966; DB 1; Length 13378;
Best Local Similarity 100.0%; Pred. No. 1,2e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACTATCTCTAAAGAGTACATGAGAGAAAAAATTTGGCAGTCCCATCACA 60
    |||||||
Db 12322 AGTAACTATCTCTAAAGAGTACATGAGAGAAAAAATTTGGCAGTCCCATCACA 12381
    |||||||

QY 61 CTATTATCATGATAGCACTTTAGCAGCTTGTTCGAAAGGTCGAGAGTGCAGACCTTATC 120
    |||||||
Db 12382 CTATTATCATGATAGCACTTTAGCAGCTTGTTCGAAAGGTCGAGAGTGCAGACCTTATC 12441
    |||||||

QY 121 AGCATGAAAGGGGATGTCATTACAGACATCAATTTATGACAGTGAAGCAACCT 180
    |||||||
Db 12442 AGCATGAAAGGGGATGTCATTACAGACATCAATTTATGACAGTGAAGCAACCT 12501
    |||||||

QY 181 TCAGCCCAACAAGTGTGTTAAATATGACATCCCAAAAGTTTGAAGAAACAATATGCG 240
    |||||||
Db 12502 TCAGCCCAACAAGTGTGTTAAATATGACATCCCAAAAGTTTGAAGAAACAATATGCG 12561
    |||||||

QY 241 TCAGAGCTTGATGATAAAGAGTTGATGATGATGATGATGATGATGATGATGATGATG 300
    |||||||
Db 12562 TCAGAGCTTGATGATAAAGAGTTGATGATGATGATGATGATGATGATGATGATGATG 12621
    |||||||

QY 301 GAAACTATCAACGCTGTCTGTGCAACAGCAGTATGACTCTTGAACACGTAACCTCA 360
    |||||||
Db 12622 GAAACTATCAACGCTGTCTGTGCAACAGCAGTATGACTCTTGAACACGTAACCTCA 12681
    |||||||

QY 361 ATTGCTACAGTAATTAATGTTAGTTGGCAGTTAAGAGTGAAGCAAGCTGAATTGACA 420
    |||||||
Db 12682 ATTGCTACAGTAATTAATGTTAGTTGGCAGTTAAGAGTGAAGCAAGCTGAATTGACA 12741
    |||||||

QY 421 GATGAAGCTTATGAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATG 480
    |||||||
Db 12742 GATGAAGCTTATGAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATG 12801
    |||||||

QY 481 CGTCTTAATATGAGATGAGCCCAAGAGTTCGAAAAAAGCCCAAGCAGAGGCTGCT 540
    |||||||
Db 12802 CGTCTTAATATGAGATGAGCCCAAGAGTTCGAAAAAAGCCCAAGCAGAGGCTGCT 12861
    |||||||

QY 541 GATTTGCTCAATTAAGCAAGATATCACTGATGATGATGATGATGATGATGATGATGATG 600
    |||||||
Db 12862 GATTTGCTCAATTAAGCAAGATATCACTGATGATGATGATGATGATGATGATGATGATG 12921
    |||||||

QY 601 GAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
    |||||||
Db 12922 GAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12981
    |||||||

QY 661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
    |||||||
Db 12982 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13041
    |||||||

QY 721 AGCCAAATTTATCATGTTGTAAGAACTCACTAAGAAACGAAAAATCATTAATTTATGATG 780
    |||||||

```

```

Db 13042 AGCCAAATTTATCATGTTGTAAGAACTCACTAAGAAACGAAAAATCATTAATTTATGATGAC 13101
    |||||||

QY 761 TACAAGAAAAATTAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
    |||||||
Db 13102 TACAAGAAAAATTAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 13161
    |||||||

QY 841 CAAGCATTTATCGAAGAAAGATTCAGACAGCAGCATATCAAGGTTAAGGACCAAGCCCTTC 900
    |||||||
Db 13162 CAAGCATTTATCGAAGAAAGATTCAGACAGCAGCATATCAAGGTTAAGGACCAAGCCCTTC 13221
    |||||||

QY 901 CAAATTAATCTTATACCATATATTCGCTGCTGAGATTCAGAGTCAAGCATAGTATCATCA 960
    |||||||
Db 13222 CAAATTAATCTTATACCATATATTCGCTGCTGAGATTCAGAGTCAAGCATAGTATCATCA 13281
    |||||||

QY 961 AACGAA 966
    |||||||
Db 13282 AACGAA 13287

RESULT 6
AX571762 349980 bp DNA linear PAT 29-NOV-2002
LOCUS AX571762
DEFINITION Sequence 4981 from Patent WO02077021.
ACCESSION AX571762
VERSION AX571762.1 GI:26003954
KEYWORDS
SOURCE
ORGANISM
    Streptococcus pneumoniae
    Streptococcus pneumoniae
    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
    Streptococcus.
REFERENCE
    1 Masignani,V., Tettelin,H. and Fraser,C.
    AUTHORS Streptococcus pneumoniae proteins and nucleic acids
    TITLE Patent: WO 02077021-A 4981 03-OCT-2002;
    JOURNAL Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
    source
    1. 349980
    /organism="Streptococcus pneumoniae"
    /mol_type="genomic DNA"
    /db_xref="taxon:1313"
    /note="seq 4979 too long: 2,162,598 bases-replaced by
    following seq:-seq 4979: from 0.000.001 to 0.349.980-seq
    4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001
    to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq
    4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001
    to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq
    4986: from 2.100.001 to 2.162,598"
BASE COUNT 108592 a 61526 c 75542 g 104320 t
ORIGIN
Query Match      100.0%; Score 966; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 5,5e-168;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACTATCTCTAAAGAGTACATGAGAGAAAAAATTTGGCAGTCCCATCACA 60
    |||||||
Db 220773 AGTAACTATCTCTAAAGAGTACATGAGAGAAAAAATTTGGCAGTCCCATCACA 220832
    |||||||

QY 61 CTATTATCATGATAGCACTTTAGCAGCTTGTTCGAAAGGTCGAGAGTGCAGACCTTATC 120
    |||||||
Db 220833 CTATTATCATGATAGCACTTTAGCAGCTTGTTCGAAAGGTCGAGAGTGCAGACCTTATC 220892
    |||||||

QY 121 AGCATGAAAGGGGATGTCATTACAGACATCAATTTATGACAGTGAAGCAACCT 180
    |||||||
Db 220893 AGCATGAAAGGGGATGTCATTACAGACATCAATTTATGACAGTGAAGCAACCT 220952
    |||||||

QY 181 TCAGCCCAACAAGTGTGTTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 240
    |||||||
Db 220953 TCAGCCCAACAAGTGTGTTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 221012
    |||||||

QY 241 TCAGAGCTTGATGATAAAGAGTTGATGATGATGATGATGATGATGATGATGATGATG 300
    |||||||

```

Db	221013	CCAGAGCGTGTGATGATTAAGAGGTTGATGATCTATTGCGGAGAGAAAAACATATATGCG	221072
Oy	301	GAAGAACTACCAACGCTGCTTGTGCACACAGAGATGACTCTTGAACACAGTAAAGCTCAA	360
Db	221073	GAAGAACTACCAACGCTGCTTGTGCACACAGAGATGACTCTTGAACACAGTAAAGCTCAA	221132
Oy	361	ATTGCTACACAGTAAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGAGTGAAGAGTGAAGTGA	420
Db	221133	ATTGCTACACAGTAAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGAGTGAAGAGTGAAGTGA	221192
Oy	421	GATGAAGCGCTATPAAGAAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATGATG	480
Db	221193	GATGAAGCGCTATPAAGAAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATGATG	221252
Oy	481	CGCTTATATATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	540
Db	221253	CGCTTATATATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	221312
Oy	541	GATTTTGGCTCAATTAAGCCAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	600
Db	221313	GATTTTGGCTCAATTAAGCCAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	221372
Oy	601	GAATTTACCTTTGATTTGCTTCAACACAGATGATGATGATGATGATGATGATGATGATGATGATG	660
Db	221373	GAATTTACCTTTGATTTGCTTCAACACAGATGATGATGATGATGATGATGATGATGATGATGATG	221432
Oy	661	GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	720
Db	221433	GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	221492
Oy	721	AGCCAAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	780
Db	221493	AGCCAAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	221552
Oy	781	TACAAAGAAAAATTTAAAAAGCTGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATG	840
Db	221553	TACAAAGAAAAATTTAAAAAGCTGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATG	221612
Oy	841	CAAGACATTAATCGAAAAAGATTTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	900
Db	221613	CAAGACATTAATCGAAAAAGATTTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	221672
Oy	901	CAAAATATCTTTACCAATATATATCGGTGGTGGAGATTCAGGCTCAAGCAGTACTACATCA	960
Db	221673	CAAAATATCTTTACCAATATATATCGGTGGTGGAGATTCAGGCTCAAGCAGTACTACATCA	221732
Oy	961	AACGAA 966	
Db	221733	AACGAA 221738	
RESULT 7	AE008462	11552 bp	DNA linear BCT 13-SEP-2001
LOCUS	AE008462		
DEFINITION	Streptococcus pneumoniae R6 section 78 of 184 of the complete genome.		
ACCESSION	AE008462	AE007317	
VERSION	AE008462.1	GI:15458482	
KEYWORDS			
SOURCE	Streptococcus pneumoniae R6		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
REFERENCE	1 (bases 1 to 11552)		
AUTHORS	Hoshino, J., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., Dehoff, B.S., Estrem, S., Filiz, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., Mcleaster, K., Mundy, C., Niclas, T.I., Norris, F.H., O'gara, M., Peck, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, R., Young-Bellido, M., Zhao, G., Zook, C., Baitz, R.H., Jaskunas, S.R., Hosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.		

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Genome of the bacterium <i>Streptococcus pneumoniae</i> strain R6	<i>J. Bacteriol.</i> 183 (19), 5709-5717 (2001)	21429245	11544234	2 (bases 1 to 11552)	Hoskins, J.A., Alborn, W. Jr., Arnold, J.D., Blaszcak, L., Burgett, S., Dehoff, B.S., Estrom, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Hann, A., Khaja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I., Norris, F.H., O'gara, M., Peery, R., Robertson, G.T., Rokey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L. and Glass, J.I.
Direct Submission	Submitted (27-JUL-2001)	Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA	Location/Qualifiers	1. 11552	
FEATURES	source	/organism="Streptococcus pneumoniae R6"	/mol_type="genomic DNA"	/strain="R6"	/db_xref="taxon:171101"
gene	62. .2436	/gene="rnr"	/note="synonym: spr0878"	82. .2436	/gene="rnr"
CDS	62. .2436	/codon_start=1	/label="spr0878"	/product="Exoribonuclease R"	/protein_id="AAK99682.1"
		/db_xref="GI:15458483"	/translation="MKRRIEYVLDQDKKVVYNDLALGKSSKDFELIKTISMERKHQRFEDGSLTEAKKHHITLKGFPHAKKNGFVSLGEEDDLFVGNKYNAI		
			DGDTVEVAVIKVADRNRKGTAAEAKI IDILHSLTVGVQVLDDEKRVAGYISKNOAI		
			KISDPYVKKSALEKTEVLAKEYIDYPSKRDIFVAVLVDVGHSTGVGIDVLEVL		
			ESMDIVSEFPAVVAEVSVDAPSQDMERDILROEITFTIDGADAKIDDAVHIK		
			ALKNGNLGVIHADVSYVYEGSALDKELNRATSYVYDRVPMPLPERLSNGICSL		
			NPSDFRLTOSALIMEIDKHGRVNYITIQVYTFKFMYSVDNIILAGDEKKREYHKE		
			IVSIEELMKAHELEMMRVKRGALNFDTEAKILYDQKGPNDIVRQGIARMEIE		
			SPHMANETVAHEHFSKIDPIYRIHEPRAEKYQKIDVASSGLATITCASISQED		
			ALQDIRAVGEPTADVLSMKLSRMOQARSNNHGHGTGLADYTIHFSPKRPFD		
			LVLHMRIDYGRSKELAEHEEOVLPETATOSSNNERAIATEREVEAMKAIEVEEYV		
			GEEEDVAVSVIVKFGVLELPEVTEGILHITNLPFPHFNERDLTLGEGSGITFRVGG		
			QOQIIRERADKMGTEIDFSEVPSEPGVIEKGLGSSRSGSGNSNRSDKEDDKRRKS		
			GRSDKRKHSQDKKKKKKPKFYEVAKKAKHGKGGKGRRT."		
gene	2399. 2866	/gene="smpb"	/note="synonym: spr0879"	2399. 2866	
CDS	2399. 2866	/codon_start=1	/label="spr0879"	/product="Small protein B"	/protein_id="AAK99683.1"
		/db_xref="GI:15458484"	/translation="MAKGECKVVAONKRRADYTIIVDTLEAGVLTGETIKSVPAARIH		
			NLKGQFQVOKGGEVWLSVNHAPYEENGINOQDEERKRLILHKLOLKEQFGKGGG		
			MTVPLKVIYIKDGXAKLLGLGAKKKHDKYDKRESIKRQDNQDINRWAKVAVNR"		
gene	2876. 3742	/gene="tens"	/note="synonym: spr0880"	2876. 3742	
CDS	2876. 3742	/gene="tens"	/codon_start=1	/label="spr0880"	/product="Tellurite resistance"
		/db_xref="GI:15458485"			

gene
/translation="MKMEKLVAKRMPLMNKQTPPEAVVQCKHNTKVGTSKIVLKG
LKFTLPEEGEVLAEHLPEAGADNPMAOPOMHREVEATDDVEVTELEPCCEPEYFK
KYNTNPVSEVLEAMOTYKOGKALDIDCGGRNSLFLAODPDVTAADONGLATEIIO
STVEDDLMDPVGIXDINSASIEDEYPISTVYVLMQLQARIATIIQNNBEKSVSG
YLVLCADIEDPCSVNPFPEKEGELADYIKDMLVKIENGGHLMRDNENRLO
LRFVTLAKKIK"

CDS
/gene="cola"
/note="synonym: spr0881"
3815..4768
/gene="cola"
/note="Competence induced protein A"
/codon_start=1
/transl_table=11
/label=spr0881
/product="competence protein"
/protein_id="AAK99685.1"
/db_xref="GI:15458486"

gene
/translation="MVARDAAGELVNVLEDKLEKQAYTCPAGSGQLHROGSPVETH
FAHSLKDCDEFEERSEPHLANKESLYHMLKKEKVOLEYLSELKOIADVPVNGNL
ALEVQCSPLPOKVLKERSEGRSOGYVLMILGQKIMLKEHLTLROGFLYFSQMGF
YVWELDKQVLRKLKLYODLRGLHYQIKEPSYGGCSLLEILRLPKKRIKISHFY
SEDKDICTIKQDLTYONLFWKEDAEATQKGENLITGLKEMTQILNPIYKGFQITE
QDLTSYQHEFTYQKNPNQNDOKLYPAFYQYFLKNME"

CDS
/gene="pepB"
/note="synonym: spr0882"
4787..6589
/gene="pepB"
/note="EC_number=3.4.-.-"
/codon_start=1
/transl_table=11
/label=spr0882
/product="Group B oligopeptidase"
/protein_id="AAK99686.1"
/db_xref="GI:15458487"

gene
/translation="WLDORNEINKEQVMDLSTIYPTDAMEBAIKDLTEQLETVAYTE
GHLDSADNLEITEFSELMERQIEKLYVYAHMKNDQTRKQOEVYAKAMTYISQL
DOAFSEPEPEMEISEKQADLEQAPLQVQYHFDKLOGKDVLSQREBELLAGA
GEIFGASAEETAILDNADIVFPYVLDGDEKVALSHGYTRIMESKKEVRYAOAL
YATVFOFHTAKTLQTNVKNQYRAKRYNKSARHALLANFVESYDNLVAAYRK
HLPLHRLERSLKTLGTSDLKMYVYPLISVEYSFTYQEAUKKAEDALAVIGEDYL
SVKRAESEERIDVYENOGKRSAGSGSYDTNAPMLNMDNDNLNLTIVHETSHM
HSSYRETPYVYGDYSIFLAEIASTVENITTEKLEVEEDDATRFAILNVEDGFR
GVFERQIOFAEHEHAIHQADQNGEVLSTDFLNKLADYADNOEYGLSKSDNEIOWE
RIPHYVYVYQYSTGFAAASALAEKIVHGSQEDRDYIDYIKAGKSDYPLNVARKA
GVDMKEDYLINDAFVFERRLNEFALVEKIGLA"

CDS
/gene="spr0883"
6591..7304
/gene="spr0883"
/note="Probable methyltransferase"
/codon_start=1
/transl_table=11
/label=spr0883
/product="Conserved hypothetical protein"
/protein_id="AAK99687.1"
/db_xref="GI:15458488"

gene
/translation="MVESYSKNANHNMRPVYKEEIVLIMROKQVTFGLKELEDF
KRNIPITIPHEVAFERFLMETMOPKNILEIGTAIGFSLALMAEAPNAKITTIDRN
EMIGFAKEAOFDSRKQTLLEGDAVVLSTLTSYDFVENDSAKSYIVLPIELK
HLEVGAVVLDITPOGDAKDIMEVRRQRTIYVGLQKLPATIDNPELITVPLG
DSILMKRVADVQLSESE"

CDS
/gene="pma"
/note="synonym: spr0884"
7371..8312
/gene="pma"
/note="EC_number=3.4.-.-"
/codon_start=1
/transl_table=11
/label=spr0884
/product="Proteinase maturation protein"
/protein_id="AAK99688.1"

gene
/db_xref="GI:15458489"
/translation="MKKKLAGATLLSVATFLAAGSKSGEADLISKGDVITEHOFY
EYVKNNSAOVUMLNIOKVEFOYSELDDEVDPTIAEEKQYGENOYRLSQAG
MTLETRKAOIRTSITVELAYKVAEALTEBAVKKADEYTPDYTAQIILNEDKAK
EVLKARAEADPAQLAKDMSPEKTEENGETLTPDASPEVEPOYKKAFAIDVDGV
SDVITATGTOYSSQITIVKTKTEKSSNIDYKREKLVILLOKONDSITFOSIIG
KELOANIKVKQDAFQNIETQYIGGSSSSSSSTNS"

CDS
/complement(8529..8771)
/gene="IS861-truncation"
/note="synonym: spr0885"
complement(8529..8771)
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label=spr0885
/product="Degenerate transposase (orf2)"
/protein_id="AAK99689.1"
/db_xref="GI:15458490"

gene
/translation="MLEQAFKEKHENTILHSDGWOYQHDYSHRFLSKGQASMR
KGNQDNGRMSEFGLKSEMFYEQTFKSLNLE"

CDS
/complement(8771..8968)
/gene="IS861-truncation"
/note="synonym: spr0886"
complement(8771..8968)
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label=spr0886
/product="Degenerate transposase (orf1)"
/protein_id="AAK99690.1"
/db_xref="GI:15458491"

gene
/translation="MYMNYTPKVRKSKSNFVGVITMKLYTDKQVQIIELRKQYSLEK
LSNKGINSNLSRYMIKILDR"

repeat_region
complement(9027..9161)

Query Match
Best Local Similarity 99.5% Score 961.2: DB 1: Length 11552;
Matches 963: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

1 AGTAACACTTATCTCAAGAGGTAGACATGAGAAAAAATTTATGGCAGGTGCATACCA 60
7344 AGTACACTTATCTCAAGAGGTAGACATGAGAAAAAATTTATGGCAGGTGCATACCA 7403

61 CTATTATAGTAGACACTTTAGCAGCTTGTTCGAAAGGTCGAAAGGTGCAGACTTATC 120
7404 CTATTATAGTAGACACTTTAGCAGCTTGTTCGAAAGGTCGAAAGGTGCAGACTTATC 7463

121 AGCATGAAAGGGAGTGCATTACAGACATCAATTTATGAGCAAGTGAAGAACCAACCT 180
7464 AGCATGAAAGGGAGTGCATTACAGACATCAATTTATGAGCAAGTGAAGAACCAACCT 7523

181 TCAGCCCAACAGCTTGTTAATATGACATCCAAAAAGTTTGAAGAAAAAATATGCG 240
7524 TCAGCCCAACAGCTTGTTAATATGACATCCAAAAAGTTTGAAGAAAAAATATGCG 7583

241 TCAGAGCTTGATGATGAAGAGAGCTTATGATCTATGCGGAGAAAGAAAAAATATGCG 300
7584 TCAGAGCTTGATGATGAAGAGAGCTTATGATCTATGCGGAGAAAGAAAAAATATGCG 7643

301 GAAACTCAACAGCTGTGTGTGACACAGAGTATGACTCTTGAACACAGTAAAGCTCAA 360
7644 GAAACTCAACAGCTGTGTGTGACACAGAGTATGACTCTTGAACACAGTAAAGCTCAA 7703

361 ATTCGTACAGTAATATGTTGAGTTGGCAGTTAAGAAAGGTGACAAAGCTGAATTGACA 420
7704 ATTCGTACAGTAATATGTTGAGTTGGCAGTTAAGAAAGGTGACAAAGCTGAATTGACA 7763

421 GATGAAGCTATAAAGAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATG 480
7764 GATGAAGCTATAAAGAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATG 7823

481 CCTCTTAATATGAGATTAAGCCCAAGAGTCTCGAAAAAGCCAAAGGCAAGAGTCTCT 540


```

Db      7824  CGCTTAATTAATGAGATAGGCCAAGAGCTTCGAAAAAGCCAGGAGAGGTCT 7883
OY      541   GATTTTGGCTCAATTAGCCAAAGATTAATCACTGATGAAAAAGAAAAGAAATGGTGA 600
Db      7884  GATTTTGGCTCAATTAGCCAAAGATTAATCACTGATGAAAAAGAAAAGAAATGGTGA 7943
OY      601   GAAATTAACCTTGAATTCGCTTCAACAGAGTACTGAGCAAGTCMAAAAGCCGCTTTC 660
Db      7944  GAAATTAACCTTGAATTCGCTTCAACAGAGTACTGAGCAAGTCMAAAAGCCGCTTTC 8003
OY      661   GCTTTAGATGATGATGCTTTCGATGATGATTAAGCAACTGCGACACAGCCCTACGT 720
Db      8004  GCTTTAGATGATGATGCTTTCGATGATGATTAAGCAACTGCGACACAGCCCTACGT 8063
OY      721   AGCCAAATTAATCAATGTTAAACTCTAAGAAAAAGAAAAATCTTAATTAATGATAC 780
Db      8064  AGCCAAATTAATCAATGTTAAACTCTAAGAAAAAGAAAAATCTTAATTAATGATAC 8123
OY      781   TACAAGAAAAATTAATAAAGCTGTTATCTGACTCAAAAACAAATGATTCACATTTGT 840
Db      8124  TACAAGAAAAATTAATAAAGCTGTTATCTGACTCAAAAACAAATGATTCACATTTGT 8183
OY      841   CAAAGCATTAATCGAAAAAGATTCGACAGCCAAATATCAAGTTAAGGACCAAGCTTC 900
Db      8184  CAAAGCATTAATCGAAAAAGATTCGACAGCCAAATATCAAGTTAAGGACCAAGCTTC 8243
OY      901   CAAAATATCTTACCCAAATATATGCTGATGATGATGATGATGATGATGATGATGAT 960
Db      8244  CAAAATATCTTACCCAAATATATGCTGATGATGATGATGATGATGATGATGATGAT 8303
OY      961   AACGAA 966
Db      8304  AACGAA 8309

RESULT 8
LOCUS   AX568726                      939 bp    DNA        linear    PAT 29-NOV-2002
DEFINITION   Sequence 1933 from Patent WO02077021.
ACCESSION   AX568726
VERSION     AX568726.1 GI:26002429
KEYWORDS
SOURCE      Streptococcus pneumoniae
ORGANISM    Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE   1
AUTHORS    Masigian, V., Tettelin, H. and Fraser, C.
TITLE      Streptococcus pneumoniae proteins and nucleic acids
JOURNAL    Patent: WO 02077021-A 1933 03-OCT-2002;
Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
SOURCE      1. 939
              Location/Qualifiers
              /organism="Streptococcus pneumoniae"
              /mol_type="genomic DNA"
              /db_xref="taxon:1313"

BASE COUNT   363 a      165 c      182 g      229 t
ORIGIN
Query Match      97.2%; Score 939; DB 6; Length 939;
Best Local Similarity 100.0%; Pred. No. 1.9e-162;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      121  CATCAATTTTATGACCAAGTGAAAAACAACCTTCAGCCCAACAAGCTCTTTAAATATG 180
OY      208  ACCATCCAAAAAGCTTTTGAATAAATATGGCTGACAGCTTGATGATTAAGAGGTGAT 267
Db      181  ACCATCCAAAAAGCTTTTGAATAAATATGGCTGACAGCTTGATGATTAAGAGGTGAT 240
OY      268  GATCACTATGCGGAGAAAAAAAACAATATGGCGAAAACTACCAAGTGTCTTGACAA 327
Db      241  GATCACTATGCGGAGAAAAAAAACAATATGGCGAAAACTACCAAGTGTCTTGACAA 300
OY      328  GCAGGTATGACTCTTGAACACAGTAAAGCTCAATTCGTACAGTAATTAATGAGTTG 387
Db      301  GCAGGTATGACTCTTGAACACAGTAAAGCTCAATTCGTACAGTAATTAATGAGTTG 360
OY      388  GCAGTTAAGAAAGGTGCGAAGCTCAATTCAGATGAAAGCTTAAATAAAGCCCTGAT 447
Db      361  GCAGTTAAGAAAGGTGCGAAGCTCAATTCAGATGAAAGCTTAAATAAAGCCCTGAT 420
OY      448  GAGTACACTCCAGATTAACGGCTCAATTCATCCGCTTAAATTAATGAAGATTAAGCCAA 507
Db      421  GAGTACACTCCAGATTAACGGCTCAATTCATCCGCTTAAATTAATGAAGATTAAGCCAA 480
OY      508  GAGTTCCTGAAAAAGCCAGGACAGAGGCTGATTTTGTCTCAATTAAGCCAAAGATAT 567
Db      481  GAGTTCCTGAAAAAGCCAGGACAGAGGCTGATTTTGTCTCAATTAAGCCAAAGATAT 540
OY      568  TCAACTGATGAAAAAACAAGAAATGGTGAGAAATTAACCTTGATTCGCTTCAACA 627
Db      541  TCAACTGATGAAAAAACAAGAAATGGTGAGAAATTAACCTTGATTCGCTTCAACA 600
OY      628  GAGTACACTGAGCAAGTCAAAAAAGCCGCTTCGCTTAAAGATGATGATGATGATGAT 687
Db      601  GAGTACACTGAGCAAGTCAAAAAAGCCGCTTCGCTTAAAGATGATGATGATGATGAT 660
OY      688  GTGATTACAGCACTGCGACACAGACCTACAGTATACCAATTAATGATTAATGATGAT 747
Db      661  GTGATTACAGCACTGCGACACAGACCTACAGTATACCAATTAATGATTAATGATGAT 720
OY      748  AAGAAACAGAAAAATCAATTAATGATGATGATGATGATGATGATGATGATGATGAT 807
Db      721  AAGAAACAGAAAAATCAATTAATGATGATGATGATGATGATGATGATGATGATGAT 780
OY      808  TTGACTCAAAAAACAATGATTCACATTTGTTCAAGCATTAATGCAAAAAAGATTCGAA 867
Db      781  TTGACTCAAAAAACAATGATTCACATTTGTTCAAGCATTAATGCAAAAAAGATTCGAA 840
OY      868  GCAGCCATATATCAAGTTAAGGACCAAGCTTCGCAAAATATCTTACCAATATATCGGT 927
Db      841  GCAGCCATATATCAAGTTAAGGACCAAGCTTCGCAAAATATCTTACCAATATATCGGT 900
OY      928  GGTGAGATTCAGGTCGAAGCAGTAGTACATCAAGCAAA 966
Db      901  GGTGAGATTCAGGTCGAAGCAGTAGTACATCAAGCAAA 939

RESULT 9
LOCUS   SPNEU1906                      11135 bp    DNA        linear    HTG 11-JUL-2001
DEFINITION   Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AL449928
VERSION     AL449928.1 GI:11545153
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae
ORGANISM    Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE   1
AUTHORS    Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Humbert, Y.,
Friedli, L., Guerrier, M., Grand-Schenck, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peltsch, M.
and Garcia-Bustos, J.F.
TITLE      Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate

```

JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21335329
 PUBMED 11442348
 REFERENCE 2 (bases 1 to 11135)
 AUTHORS Dopezo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A., Humbert, Y., Friedl, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Buehl, G., Feger, G., Garcia, E., Feltsch, M. and Garcia-Bustos, J. F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..11135
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"
 BASE COUNT 32875 a 20454 c 25072 g 32680 t 54 others
 ORIGIN
 Query Match 97.2%; Score 939; DB 2; Length 11135;
 Best Local Similarity 99.5%; Pred. No. 6.4e-163;
 Matches 942; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 AGTAACTATCTCAAGAGAGATAGACATGAAAGAAAATTTGGCAGCTCCATCACA 60
 Db 110189 AGTAACTATCTCAAGAGAGATAGACATGAAAGAAAATTTGGCAGCTCCATCACA 110248
 Oy 61 CTATTATCAGTACGACCTTATGACCTGTTGGAAGGCTGGAAGGTGACAGCTTATC 120
 Db 110249 CTATTATCAGTACGACCTTATGACCTGTTGGAAGGCTGGAAGGTGACAGCTTATC 110308
 Oy 121 AGCATGAAAGGGAGTGTCTATTACAGAACATCAATTTATGACAGTGAAGCAACCT 180
 Db 110309 AGCATGAAAGGGAGTGTCTATTACAGAACATCAATTTATGACAGTGAAGCAACCT 110368
 Oy 181 TCAGCCCAACAGCTGTGTTAATATGACATCCAAAAGTTTGAAGAAAACAATATGCG 240
 Db 110369 TCAGCCCAACAGCTGTGTTAATATGACATCCAAAAGTTTGAAGAAAACAATATGCG 110428
 Oy 241 TCAGAGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 Db 110429 TCAGAGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110488
 Oy 301 GAAAGCTACAGCTGTCTGTCACAGAGGATGATGATGATGATGATGATGATGATGATG 360
 Db 110489 GAAAGCTACAGCTGTCTGTCACAGAGGATGATGATGATGATGATGATGATGATG 110548
 Oy 361 ATTCTACAGTAAATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTA 420
 Db 110549 ATTCTACAGTAAATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTA 110608
 Oy 421 GATGAAGCTATTAAGAAAGCCCTTTGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 110609 GATGAAGCTATTAAGAAAGCCCTTTGATGATGATGATGATGATGATGATGATGATGAT 110668
 Oy 481 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCTCGAAAAAGCCAGGCAAGAGTGTCT 540
 Db 110669 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCTCGAAAAAGCCAGGCAAGAGTGTCT 110728
 Oy 541 GATTTGCTCAATTAAGCCCAAGATTAATCACTGATGATGATGATGATGATGATGATGAT 600
 Db 110729 GATTTGCTCAATTAAGCCCAAGATTAATCACTGATGATGATGATGATGATGATGATGAT 110788
 Oy 601 GAAATTAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 Db 110789 GAAATTAACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110848

Oy 661 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
 Db 110849 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110908
 Oy 721 AGCCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
 Db 110909 AGCCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 110968
 Oy 781 TACAAAGAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
 Db 110969 TACAAAGAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 111028
 Oy 841 CAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
 Db 111029 CAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 111088
 Oy 901 CAAATATCTTTTACCAATATATATGATGATGATGATGATGATGATGATGATGATGAT 947
 Db 111089 CAAATATCTTTTACCAATATATATGATGATGATGATGATGATGATGATGATGATGAT 111135
 RESULT 10
 AR120254
 LOCUS AR120254 640 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 33 from patent US 6159469.
 ACCESSION AR120254
 VERSION AR120254.1 GI:14103830
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Chon, G. H., Kunsch, C. A., Barash, S. C., Dillon, P. J., Dougherty, B., Fannon, M. R. and Rosen, C. A.
 JOURNAL Streptococcus pneumoniae antigens and vaccines
 Patent: US 6159469-A 33 12-DEC-2000;
 FEATURES
 source
 1..640
 /organism="unknown"
 BASE COUNT 234 a 112 c 141 g 153 t
 ORIGIN
 Query Match 57.2%; Score 552.4; DB 6; Length 640;
 Best Local Similarity 94.7%; Pred. No. 1.7e-91;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;
 Oy 90 TTCGAAGGGTGAAGAGGTCAGACCTTATGACATGAAAGGGATGTCATTACAGAAACA 149
 Db 1 TTCGAAGGGTGAAGAGGTCAGACCTTATGACATGAAAGGGATGTCATTACAGAAACA 60
 Oy 150 TCAATTTATGAGCAAGTGAAGAAAGCAACCTTCAGCCCAACAGTCTTTGTTAATATGAC 209
 Db 61 TCAATTTATGAGCAAGTGAAGAAAGCAACCTTCAGCCCAACAGTCTTTGTTAATATGAC 120
 Oy 210 CATCCAAAAGTTTGAAGAAACATATGCTCAGAGCTTGATGATTAAGAGAGCTGATGA 269
 Db 121 CATCCAAAAGTTTGAAGAAACATATGCTCAGAGCTTGATGATTAAGAGAGCTGATGA 180
 Oy 270 TACTATGCGGAAGAAAACCAATATGCGGCAAACTACAGACGCTTGTCACAAAC 329
 Db 181 TACTATGCGGCAAGAAAACCAATATGCGGCAAACTACAGACGCTTGTCACAAAC 240
 Oy 330 AGGTATGACTTGTGAACACAGTAAAGCTCAAAATGCTATACAAATTAATTAATTAATTAAT 389
 Db 241 AGGTATGACTTGTGAACACAGTAAAGCTCAAAATGCTATACAAATTAATTAATTAATTAAT 300
 Oy 390 AGTTAAGAGGTAGCAAGAGCTGAATGACACATGAAGCTTTAAGAAAGCTTTGATGA 449
 Db 301 AGTTAAGAGGTAGCAAGAGCTGAATGACACATGAAGCTTTAAGAAAGCTTTGATGA 360
 Oy 450 GTACACTCAGATGTAAGGCTCAATCATCGCTTAATTAATTAATTAATTAATTAATTAAT 509
 Db 361 GTACACTCAGATGTAAGGCTCAATCATCGCTTAATTAATTAATTAATTAATTAATTAAT 420

QY 510 AGTTCGAAAAAGCCAGAGAGTGTCTATTTGCTCAATTAGCCAAAGATATTC 569
|||||
DB 421 AGTTCTCAAAAAGCCAGAGAGTGTCTATTTGCTCAATTAGCCAAAGATATTC 480
QY 570 AACTGATGAAAAACAAAAAATGTGTGAGAAATTAACCTTTGATTCGTTCAACAGA 629
|||||
DB 481 AACTGATGAAAAACAAAAAATGTGTGAGAAATTAACCTTTGATTCGTTCAACAGA 540
QY 630 AGTACCTG--AGCAAGTCAAAAAGCCGTTTCGTTAGATGTGATGCGTGTCTGAT 687
|||||
DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGTTTCGTTTCGTTAGATGTGATGCGTGTCT 600
QY 688 G-----TGATTACAGCAACTGGCAGACA 710
|||||
DB 601 GGATGTGATTACAGCAACTGGGCGACA 628

RESULT 11
BD063263 640 bp DNA linear PAT 27-AUG-2002
LOCUS Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063263.1 GI:22608866
VERSION JP 2001505415-A/17.
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 640)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromocky,J.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 17 24-APR-2001;
HUMAN GENOME SCIENCES INC
PN JP 2001505415-A/17
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,SYDOR L JOHNSON,ALEX HROMOCKYJ PC
C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source 1..640
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 234 a 112 c 141 g 153 t
ORIGIN

Query Match 57.2%; Score 552.4; DB 6; Length 640;
Best Local Similarity 94.7%; Pred. No. 1.7e-91;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTCGAAAGGGTCAGAGTGCAGACCTTTCAGCATGAAGGGATGTCATACAGAA 149
|||||
DB 1 TTCGAAAGGGTCAGAGTGCAGACCTTTCAGCATGAAGGGATGTCATACAGAA 60
QY 150 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAAGTCTTGTAAATATGAC 209
|||||
DB 61 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAAGTCTTGTAAATATGAC 120
QY 210 CATCCAAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTGATGA 269
|||||
DB 121 CATCCAAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTGATGA 180
QY 270 TACTATGCGGAGAAAAACAATATGCGAAAACTACCAACGTCCTTGTCACAAAGC 329
|||||
DB 181 TACTATGCGGAGAAAAACAATATGCGAAAACTACCAACGTCCTTGTCACAAAGC 240
QY 330 AGGTATGACTTGTGAACAAGTCAAAATTCGTACAGTAATTAAGTTGAGTTGAC 389

DB 241 AGGTATGACTTGTGAACAAGTCAAAATTCGTACAGTAATTAAGTTGAGTTGAC 300
|||||
QY 390 ACTTAAGAAAGTACAGAAAGCTGATTTGACATATGAAGCTTTAAGAAAGCTTTGATGA 449
|||||
DB 301 ACTTAAGAAAGTACAGAAAGCTGATTTGACATATGAAGCTTTAAGAAAGCTTTGATGA 360
QY 450 GTACACTCCAGATGTACAGGCTCAATCATCCGTTTATTAATGAAGTAAAGCCCAAGA 509
|||||
DB 361 GTACACTCCAGATGTACAGGCTCAATCATCCGTTTATTAATGAAGTAAAGCCCAAGA 420
QY 510 ACTTCTGAAAAAGCCAGAGGAGTGTGATTTGCTCAATTAGCCAAAGATATTC 569
|||||
DB 421 ACTTCTGAAAAAGCCAGAGGAGTGTGATTTGCTCAATTAGCCAAAGATATTC 480
QY 570 AACTGATGAAAAACAAAAAATGTGTGAGAAATTAACCTTTGATTCGTTCAACAGA 629
|||||
DB 481 AACTGATGAAAAACAAAAAATGTGTGAGAAATTAACCTTTGATTCGTTCAACAGA 540
QY 630 AGTACCTG--AGCAAGTCAAAAAGCCGTTTCGTTAGATGTGATGCGTGTCTGAT 687
|||||
DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGTTTCGTTTCGTTAGATGTGATGCGTGTCT 600
QY 688 G-----TGATTACAGCAACTGGCAGACA 710
|||||
DB 601 GGATGTGATTACAGCAACTGGGCGACA 628

RESULT 12
ARI93967 791 bp DNA linear PAT 20-APR-2002
LOCUS ARI93967
DEFINITION Sequence 219 from patent US 6348328.
ACCESSION ARI93967
VERSION ARI93967.1 GI:20240559
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 791)
AUTHORS Black,M.Terence., Hodgson,J.Edward., Knowles,D.Justin,Charles.,
Nicholas,R.Oakley. and Stodola,R.King.
TITLE Compounds
JOURNAL Patent: US 6348328-A 219 19-FEB-2002;
FEATURES
source 1..791
/organism="unknown"

BASE COUNT 258 a 140 c 168 g 224 t 1 others
ORIGIN

Query Match 27.5%; Score 265.6; DB 6; Length 791;
Best Local Similarity 97.9%; Pred. No. 6.7e-39;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTAATGCGAGGTCCATCACA 60
|||||
DB 507 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTAATGCGAGGTCCATCACA 566
QY 61 CTATTATCATGACCACTTTAGCAGCTTGTGTGAAGGGTCAGAAAGTGTCAACCTTATC 120
|||||
DB 567 CTATTATCATGACCACTTTAGCAGCTTGTGTGAAGGGTCAGAAAGTGTCAACCTTATC 626
QY 121 AGCATGAAGGGGATGTCATTACAGAAACATCAATTTATGAGCAAGTGAAGAACCTTATC 180
|||||
DB 627 AGCATGAAGGGGATGTCATTACAGAAACATCAATTTATGAGCAAGTGAAGAACCTTATC 686
QY 181 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGTG-AAAAACAATATG 239
|||||
DB 687 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGTG-AAAAACAATATG 746
QY 240 CTCAGAGCTTGATGATTAAGAGGTGATGATCTATATGCGGAGAGA 284
|||||
DB 747 CTCAGAGCTTGATGATTAAGAGGTGATGATCTATATGCGGAGAGA 791

```

RESULT 13
LOCUS       AE014908                      12012 bp    DNA        linear    BCF 25-Oct-2002
DEFINITION  Streptococcus mutans UA159 section 56 of 185 of the complete genome.
ACCESSION   AE014908 AE014133
VERSION     AE014908.1 GI:24377016
KEYWORDS
SOURCE      Streptococcus mutans UA159
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 12012)
AUTHORS     Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,
            Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,
            Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and
            Ferretti,J.J.
            Genome sequence of Streptococcus mutans UA159, a cariogenic dental
            pathogen
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14434-14439 (2002)
MEDLINE     22295063
PUBMED      12397186
REFERENCE   2 (bases 1 to 12012)
AUTHORS     Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,
            Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,
            Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and
            Ferretti,J.J.
            Direct Submission
TITLE       Submitted (09-JUL-2002) Department of Microbiology and Immunology,
            University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
            Oklahoma City, OK 73104, USA
FEATURES
            source
            location/Qualifiers
                1..12012
                /organism="Streptococcus mutans UA159"
                /mol_type="genomic DNA"
                /strain="UA159"
                /db_xref="taxon:210007"
                95..1063
                /locus_tag="SMU.641"
                95..1063
                /locus_tag="SMU.641"
                /note="Best Blastp Hit: emb|CAB65629.1| (A1136149)
                putative oxidoreductase. [Streptomyces coelicolor A3(2)]"
                /codon_start=1
                /transl_table=11
                /product="putative oxidoreductase"
                /protein_id="AAN58375.1"
                /db_xref="GI:24377017"
                /translation="MKAVYIOKGTGGPATIIVGSLVPPKVRPKSVLIKVAIASINYYDL
                FRSGIQTDLNPPIILGRDAIGVVEIGKDVIOFRVGCWNTNSMGVDRGQITSEY
                ALIPERLELTIRKADSLKLGAVSAATATVLODIMELKPKOKLIEGAAGHGSK
                LVLNEMGAEVVTVASLKDPSRLKOLSDPAICFDYDKSLFKLSAAYDGFPHIIDT
                SEEILQECDLAKRGVVISLITAPKDSQDSRQVYVNMCOOKGFVYIASHAKOLKKA
                GRLINDAFROGOLLEEDVYIKFKDAAEHDMETKKEKRKILIVP"
                1308..1589
                /locus_tag="SMU.642"
                1308..1589
                /locus_tag="SMU.642"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAN58376.1"
                /db_xref="GI:24377018"
                /translation="NNILDRYVSLDLGAKRKRKLTCLCNFSENAITEANDGNTYSG
                REINPFEKFKNSSELKHLMDTKELIADDMRQIGRLVAAKLESTLL"
                1695..2588
                /locus_tag="SMU.643"
                1695..2588
                /locus_tag="SMU.643"
                /note="Best Blastp Hit: pf11718306A xylanase
                [Butyrivibrio fibrisolvens]"
                /codon_start=1
                /transl_table=11
                /product="putative esterase"
                /protein_id="AAN58377.1"
                /db_xref="GI:24377019"
                /translation="MKINTSMLEWEDKAKSEVYTYTYLIEKTPDINKNRPRAVIICGG
                GGEFKYDREKEPEVALYFLNKGFOATLTDYIVGKSGYKRPPLYDKMLLYIRENNON
                WNIDETKVIYIGFSAGATHASASLANSMNPLOKYPDYGETLKPSELVLLSYPLDEN
                VYEGVSDPDNOMPTKLSPMPRIDELVTGALKTVYGVKELTETLKEKSPTEHSTATV
                PTFIEMODDDCIYNNALLEYAKRLKDNILVLEHMFALGEHGLSYINSNTQAKPSDY
                FELGTRKICICQFINRVHLK"
                2896..3834
                /locus_tag="SMU.644"
                2896..3834
                /locus_tag="SMU.644"
                /note="Best Blastp Hit: gb|AAC3746.1| (AF052209)
                competence protein [Streptococcus pneumoniae]"
                /codon_start=1
                /transl_table=11
                /product="putative competence protein/transcription
                factor"
                /protein_id="AAN58378.1"
                /db_xref="GI:24377020"
                /translation="MLVARDKNGKLLNLNGIPDKSDPFCPCACQSEVRLKNGWRPH
                FAVALQDCKEYSENESAEHLNKAELVOSLSQTESVEIEKVIPEEQTADLLVHNLT
                ALEVOCISRLSEARLCEKERTQAYQANGOVLMLEKELIMDLRSLNHLKQFLYSQNGF
                HLEWELDINRREIRLKYLIYEDIFGKYYQYFRSCSEPDGNTMTFLOLPYAKOTLYSPVH
                ORROVGLAIQKOLARPNMLRQOELAYSQGRNLLIADSDAFPPQVRLPQCSRGFCOI
                NODVSHFSAAFFQYQKQKKEIQTLVPPAFYDKMK"
                3856..5655
                /gene="pepB"
                /locus_tag="SMU.645"
                3856..5655
                /gene="pepB"
                /locus_tag="SMU.645"
                /note="Best Blastp Hit: sp|O53778|PEPB_SRRAG GROUP B
                OUTGOPEPTIDASE PEPB>g1112786731|p11751748 chimet
                oligopeptidase (EC 3.4.24.15) pepB [validated] -
                Streptococcus agalactiae>g111469939|gb|AAC44215.1|
                (U49821) group B oligopeptidase Peps [Streptococcus
                agalactiae]"
                /codon_start=1
                /transl_table=11
                /product="putative oligopeptidase"
                /protein_id="AAN58379.1"
                /db_xref="GI:24377021"
                /translation="MSDNRSHTIEEKYQMDLTSIPPDQWKEEKSIAKDLTAIVKRA
                GILDSOSLLEITELXYLASRLREKLYVYASMKNDQDTVSYQEOLOKKAATLAKF
                NOTPAFYEPPEMLISTEKYQDPLADCPDLPLPIQHPFEKLFKKHVLKSKERALLAE
                SEIFGAGETFEELIDNADILIPFVAVDTEEEVELSHGNFISIMESKNRRVROEAVKAM
                VSIIEYQHTVAKTLDQTNVKNVNFQAKIRNYVSARBAALANFIPETVYDILVEAVHN
                HPLDLORYKTVKLTQGLDKMYDLYVLESDMTFYEEALDLAAEVLAIYFGROYA
                EHVHNAFRERMDIVRVNKNKRSAGSGSYDYNNAFLIMQOGLDLDFLVLVEGTSTY
                HSTFIRENPQYVYGGDISIFLAELASTNTNNTLTKATDEKORFATLNYLYDGFK
                GTVFROTOFAFEEDDILYKADQAGCVLTSYNNILNVAADLIEKTYGKSKENNPLOTEKA
                RIPHRYNRYVYQYATGFAASFLAKHYHGNQEDDKLTNTLAKGNSDYPVLNAYKA
                GUDWTESTYLDQAFVFEERLLELEVEKGID"
                5868..6485
                /locus_tag="SMU.646"
                5868..6485
                /locus_tag="SMU.646"
                /note="Best Blastp Hit: p111A57362 gyrB protein -
                Streptococcus pneumoniae (fragment)
                >g11052803|emb|CA58770.1| (X83917) orf1gyrb
                [Streptococcus pneumoniae]>g111490396|emb|CAA91552.1|
                (267740) unidentified [Streptococcus pneumoniae]"
                /codon_start=1
                /transl_table=11
                /product="putative phosphatase"
                /protein_id="AAN58380.1"
                /db_xref="GI:24377022"
                /translation="MPQAFIMWLDIGTLDSYEAIMEALAVYTRTFGPFQFAPIRY
                IIKESVGKLDILAKOHGISSEELKTFPIQEOYKRDKDKIKLMPFAKELQWAKDKGKIK

```

	gene		YFGRDRIADVEAENAGIKSLINLQPOSAANQKIDNLVQLQKKP"
CDS		/locus_tag-"SMU_647"	6535..7242
		/locus_tag-"SMU_647"	6535..7242
		/note-"Best Blastp Hlt:	emb CAA68045.1 (X99710)
		methyltransferase [lactococcus lactis]"	
		/codon_start-1	
		/transl_table-1	
		/product-"putative methyltransferase"	
		/protein_id-"AA58381.1"	
		/db_xref-"GI:24377023"	
		/translation-"MKSYSKNANHNRPRVIKDIIHYMRHEOKENNGLKELEIIRA	
		HENIRPIIQETVVVFRLQTQPKNILEGTAIGFSALMANAPQANTTTIDRNP	
		EMLIAKENPAKYSDRSKOINLEDGANILTKLGANDFEWDSAKSKEYIVELPTVEE	
		CLETGIENVLDVDFOGGDIAPKPIEVHQRGTIVRGLOKLFDATLKRPDLTVLSLPDG	
		DGILLIKRKREICIIDIP"	
gene		7304..8305	
		/gene_"prtct"	
CDS		/locus_tag-"SMU_648"	7304..8305
		/gene_"prtct"	
		/locus_tag-"SMU_648"	
		/note-"Best Blastp Hlt:	spl OQ2473 PRTM_LACRA PROTEASE
		MATURATION PROTEIN PRECURSOR >gll1348576 pir AA4858	
		probable protein export protein prtm precursor -	
		lactobacillus paracasei (strain NCD0_151)	
		>gll149581 gb JAA525247.1 (M83946) maturation protein	
		[lactobacillus paracasei]"	
		/codon_start-1	
		/transl_table-1	
		/product-"putative protease maturation protein precursor"	
		/protein_id-"AA58382.1"	
		/db_xref-"GI:24377024"	
		/translation-"MKRRITATGVLTLLSTYLTLACSCTNSTONSKIATMGKITVADEP	
		TNEKNSIASQAVALSVLRKFPEFKQGDKAYSDSEVTIKATEAKTIGEDSSVLAS	
		GYYEKARVKQDIRSELKLEYAVEAKEAEKIETDAISAKDVREPVTAVQILOJSEDKKA	
		KSYLEEARKADGADEPAKTIANNTKGDTIEYSFGSGSTNIPLSOVLISAUNILDGDVSDF	
		KASTTIYKPVYIVKTYIKTKTDKNADMWKARYKRKEITIISOGLKDNSRNNAVICKAF	
		RANSKINDKAFSEILSQYAASGSSSGSTTTTTNASASAATTADDQOTLAETTAAE	
'\t'gene		8544..9044	
CDS		/locus_tag-"SMU_649"	8544..9044
		/locus_tag-"SMU_649"	
		/note-"Best Blastp Hlt:	pir G70079 hypothetical protein
		yxj1 - Bacillus subtilis >gll1783233 dbj BBA11710.1	
		(DB3026) hypothetical [bacillus subtilis]	
		>gll2636429 emb CAB15920.1 (Z99123) yxj1 [bacillus	
		subtilis]"	
		/codon_start-1	
		/transl_table-1	
		/product-"conserved hypothetical protein"	
		/protein_id-"AA58383.1"	
		/db_xref-"GI:24377025"	
		/translation-"MSFOIKOKMMSPGCKFTTIDEDELPIPYOVEGSIIFKIPKTFIITS	
		DMGQKQSRIORKTWFTELPPRYVNLMHDOSFFFLKRDLSFRPHYTINKOLDMEIOGFWE	
Query Match	27.5%;	Score 265.4;	DB 1; Length 12012;
Blast Local Similarity	57.6%;	Freq. No.	3.9e-39;
Matches 504;	Conservative 0;	Mismatches 356;	Indels 15; Gaps 14;
Oy	47	CAGGTGCATCACACTATTATCAGTAGCAAGCATTTAGCGTCCTTGTCGAAGAAGGCCAGAAC	106
Db	7326	CAGAGACTTGTTACTCTTTCGTATGTCATTTAGCGGCTTGTCCAAAACAACCANA	7388
Oy	107	GTGCAGACCTTATCAGCATGGAAGGGATGTCAATTACAGAACATCAATTTATGAGCAAG	166
Db	7386	ACAGTAAATAATTGTACATCAATGAAGGCGTACTACTATTACTGTTGCAGATTTTATATATGAAG	7445
Oy	167	TGAAAAAGCAACCCCTTACAGCCAACAAGCTCTGTTAAATATGACCATCCAAAAGTTTGTG	226
Db	7446	TAAAGAAATTCAACAGCTTCTTAAAGCAGCGCTGATTATATCTCTTAAATTCCAAGTTTGTG	7505

[illegible]


```

/translation="MEKTKRYIIATAGILHLIMLGSTYAMSVENPNILOETGMDQAPV
AFAPSLFCLGLSAFAMGNLVEQYRPLTSTVSAIYASNMLTGLAIDKEIEMLY
IGVIGIGLGGAGYITPISITIKMPDKRMATGPAIMSGFSLTSPLOKLIET
EGVATPFIYLGILYIMLPASOLIKPTAEIADIDKRIKONNSYILEGTAKALK
TKSFYCLMVLIFINITGGLISVAPPAODTGMSPMSIIVGAMGIFGFRVW
ASLSDYIGRRTVILFLVSLIMTISLIFAMSLIFMISITMTVCYAGSLIPYL
SDLEGAKEIATLHGYILTAWAIALTGTMLSTIVETHNLTLCVEFVILYIIGLW
ALRLKK"
gene
complement(8593..10398)
/feature="pep"
/note="synonym: SPY1393"
complement(8593..10398)
/feature="pep"
/EC-number="3.4.24.15"
/note="Best Blastp hit = sp|P54124|PERL_LACIA
OLIGOENDOPEPTIDASE F, PLASMID >gi|2126618|pir|I.A5485
11goendopeptidase F - Lactococcus lactis
>gi|510140|emb|CAAB354.1| (23252) 11goendopeptidase F
[Lactococcus lactis]"
/codon_start=1
/transl_table=1
/product="putative oligopeptidase"
/protein_id="AAK34212.1"
/db_xref="GI:13622496"
/translation="MTDNRSILREKYTMDLSTIPATDKMWEAEVSDIATEVASKGEA
GHLDSSANLKVTKYELARVEKYVYAHMNDOPITVAKYOEOAKASGLYAKF
SEVSPFDEYVMLEHEDYQALFETPELKYNHPPKLFQARFHVLSQAEELIAGA
QTEINGAEFESTLNDIYPPVYKNDKGEVETLHGFIISMSKDSRVQRAYEAM
KSTYEQPHYAKTLQTNVQYQKARHKDKSROAMAMNFIPEAYITDLELVNK
HPLHRYLKLROVELGLDLMKYDVTPLSETDLAIGDALKAEVILAVFEKDYA
DVRARFETRMIDIVHNKGRSGVSGSYDPTNFILNMDOTLNDLTLLHGHSL
HSTPREQPYVDGYDSIFLAISTETMENLAEVODEKERERETLILHYDGHSL
GTVRQTOFAFEFHAHQADQGEVLSEVYNOLYADINEKYVGLSKDNFIQYEMA
Query Match 27.1% Score 261.6; DB 1; Length 11615;
Best Local Similarity 56.9%; Pred. No. 2e-38;
Matches 505; Conservative 0; Mismatches 374; Indels 9; Gaps 1;
34 AAAAAATATATGGAGGTGCATATATATGACAGCACTTATGACAGCTTGTCG 93
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6173 AATTAACATCTGCTAGTGTGTGACATTTGGCCCTGACGATGCTTTAGCAGCTTGCAA 6114
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 AAAGGCTCAGAGGTGCACCTTATGACATGAAGGAGTGTCTATTAAGCAATCAAA 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6113 TCACTAATGACATTAAGTATATTCGATGAAGGTGATTAAGGTTATGATGAT 6054
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 TTTTATGAGCAAGGAAAGCAACCTTCAGCCCAACAAGTCTTGTAAATATGACCTC 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6053 TTTTACATGAACAAAAACAGAAATATCGCAAAAAGGAGTCTTAATCTGTTAAT 5994
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 CAAAAAGTTTGAATAACATATGAGCTCAGAGCTTGATGATTAAGAGGTGATGATCT 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5993 AGTGTGTTTGAAGCTCAATATGTGATTAAGGTTTCAAAAAAGAGCTTGAAGAGCG 5934
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 ATTGCCGAAGAAAAAACAATATGCGCAAAACTACCAACGTGCTTCTCAACAGCAGT 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5933 TATCATATAACAGCTGACAGATATGCGCTCATCTCTGCTGCTTGACCAATCAAGC 5874
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334 ATGACTCTGAAACAGCTTAAGCTCAATTTGTACAACTAATTTAGTTGAGTTGCCAGT 393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5873 TTGACACTGAGACTTTTAAAGCTCAGATCGCTCTTCAAAAATAGTAAATATGCGGTT 5814
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 AAGAAGGTAGAGAAAGTGAATTTGACATGAGAGCTTAAAGAAAGCTTGTGATGATGATC 453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5813 AAAGAGAGAGCTTAAAAAAGATTTGACACACAGATATTAAGAAAGCATATGAATCTTAT 5754
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 ACTCCAGATGTAAGCGCTCAATATCGCTTAAATTAATGAAGTAAAGCCCAAGAGATT 513
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5753 ACTCCAAATATGCGAGCTGATGATTTAGTAAATGAAGACAGCTAAATCAAGTC 5694
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 CTCGAAAAAGCCAGGAGAGGTGCTGATTTTGTCTCAATTTAGCCAAAGATTAATCAACT 573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5693 TTAGAGAACTTAAAAAGCGCAGAGACTTTTACAGCTATTTGCTAAAGAAAAAACAAAC - 5635

```

```

OY 574 GATGAAAAACAAAAAGAAATATGTCGAGAAATATCTTGTATGCTTCAACAGACTA 633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5634 -----AACACCTGAGAAAAAAGATGACCTATTAATTTGATTCAGGTCGACAAATGTA 5583
OY 634 CCGTAGCAAGTCAAAAAAGCCGCTTCCCTTTAGATGATGATGATGATGATGATGAT 693
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5582 CGAGATGATGTCGTAAAAAGCGGCTTCAAGTTTGAATGAGGGTGGCATATCAGAGCTTATC 5523
OY 694 ACAGCACTGCGACACCAAGCCTACAGTACGCAATTTTCAATTTTAAACTACTAAGAAA 753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5522 TCGGTTTAAATCCAACTCTTATCCAAAGAGTTTACATTTGTTAAGTCACTAAAAA 5463
OY 754 ACAGAAAAATCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5462 GCAGAAAAAATTCAGATTTGCGACAGATATTAAGAAAGCTTTGAAGATATATTTAGCT 5403
OY 814 CAAAAACAAAAATGATTAACATTTGTTTCAAGCATTTATCGAAAAAGAAATTCAGAACAGCC 873
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5402 GAAAAATCAAAAAAGATATGATTAATTTCCAAAAAGATTTATGCAAAATGATGATGATTAAGCT 5343
OY 874 AATATCAAGTTAAGGACCAAGCCTTCCAAATATCTTTACCAATAT 921
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5342 AATGTAATAATTAAGACAAAGCTTTGCTAATATTTTGGCGCAATAT 5295
RESULT 15
AE014157 51454 bp DNA linear BCT 19-JUL-2002
LOCUS Streptococcus pyogenes MGAS315, section 22 of 37 of the complete
DEFINITION genome.
ACCESSION AE014157 AE014074
VERSION AE014157.1 GI:21904768
KEYWORDS Streptococcus pyogenes MGAS315
SOURCE Streptococcus pyogenes MGAS315
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Genome sequence of a serotype M3 strain of group A Streptococcus:
phage-encoded toxins, the high-virulence phenotype, and clone
emergence
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
2 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Direct Submission
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
FEATURES
source
1. 51454
/organism="Streptococcus pyogenes MGAS315"
/mol_type="genomic DNA"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/note="group: A"
complement(150..1139)
/feature="grab"
/feature="grab"
/note="synonym: SPYM3_1032"
complement(150..1139)
/feature="grab"
/note="best blastp hit: gb|AA026342.1|AF124403_1
(AF124403) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=1

```

```

/product-"protein G-related alpha 2w-bindingprotein"
/protein_id-"AAM79639.1"
/db_xref="GI:21904769"
/translation-"MGKEIKRYFLRBSAFGLVAASVIVSGTVSAVSPPEPRRI
DALOSEAAVVSOTDASDADLEALADQTDALQSEAAVYKADNADSALEALADTDA
VQSEAAVYKADNADSALEALADQTDALQSEAAVYKADNADSALEALADTDAQ
SEEAADVADNADSALEALADQTDALQSEAAVYKADNADSALEALADTDAQ
TKDTPVYKKEERQVNTLPPTGESNPEFTAAALAIWSTGALVSSCKEN"
complement(1496..2710)
/gene="murZ"
/note="synonym: Spym3_1033"
complement(1496..2710)
/gene="murZ"
/note="best blastp hit: gb|AAL97966.1| (AE010057) putative
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product-"putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id-"AAM79640.1"
/db_xref="GI:21904770"
/translation-"MAVSGAKNSVALIPAILADYILDGVPALISVDLSIETMEL
MGATVNHGDTLEDPRGVODIPMYGKINSIRASVYRYSGLGRFGOAVGLPGCD
LGPRIDILHLAEAMGVSEEGENMLSTNGKIHGAHIMDVSGATINTVAA
TKAOGTILVEMNAREPEIDVATLNNMGAIHRCAGTDITTOGVOKLHGRHYIPD
RIEAGTYIALAAIGKVTINVLVLEHSEFIAKLEMGVRMTVEDAIEVEKOSKL
ATTIKTSYPGPATDLOOPLMPLIKADGRGIIIDTYEKRINHYPBELMRGADISVI
GGQIYVQSPRLTGQVAKATDRAGAAVYTAGLMAEGKTEITINIEFIIRGASITAKL
TALGADLIOLID"
complement(2929..4125)
/gene="mekK"
/note="synonym: Spym3_1034"
complement(2929..4125)
/gene="mekK"
/note="best blastp hit: gb|AAL97967.1| (AE010057)
S-adenosylmethionine synthetase [Streptococcus pyogenes
MGAS8232]"
/codon_start=1
/transl_table=11
/product-"putative S-adenosylmethionine synthetase"
/protein_id-"AAM79641.1"
/db_xref="GI:21904771"
/translation-"MSERKLTSESSEGHDPKIDAOISDAIILDAIILDEBEAHVAE
TCVYVSGVHVEGEISTAYIDINRVVRDTIAIGTEAEYGSASVGVHSLVQSG
DIAOGVNALSRBEDTDLSHIGADQGLMEGAINTPELMPPISTSHQVLRILA
ELRSGEISYLRPDAKSOVYVEYDEHDKPVRVDVYSTOHDPENAOIRODVTEKY
IKAVIPADYIDDDPKEFINPTGRPIYVGGPOGSGLTGKTIIVDTYGGSRGGAFSG
KDATKVDASATYARTIKNTLVAGLATKAEOVLATAGVADPVSVRVDFTGISTVPE
AVLEAAVRQVDFLRPAGIITOMLDRKRPYIKOTPAAYAGHGRDIDILPWEHLKVDALVE
AVKTVL"
complement(4662..7040)
/gene="inla"
/note="synonym: Spym3_1035"
complement(4662..7040)
/gene="inla"
/note="best blastp hit: gb|AAK34188.1| (AE006574) putative
internalin A precursor [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product-"putative internalin A precursor"
/protein_id-"AAM79642.1"
/db_xref="GI:21904772"
/translation-"MKTKVILVGLLSQTLTLCOSRGNGTYPIKTKOSRKGMTS
NKIRPKSKTKNTKNGVAGVDPTDGGFILTCKSLKSTDOGIIVDHGHSJIF
YADKSGPEYLIPKASLAPVAAQQAASOGTSKVDAPHHHYEPNADIVAEALIG
TVRHDHRYILKSLSSGTOQAQAVATRLPQTSPTSTANCTPCLHPPTSGFRQ
FNGGIVGVTKDNILVHDGHDHPISTFSDRQGGAAVADQIDEPKAKAEPAETHTP
ELSEREKYOEKLAIVIAEKIDPSTIKRVTQSGKGLGEPHHDHAAVLMISDEIG
KDIIDPHAIIEHARELEKHFQMDTLRALGDEEVLIDIVTIDADPTPPSEKEDNM
KEMLATVILKIDGSRKDPLOKRGSLPNELEVIGGFTPIDISPVLOFKKIKOLIMT
KTGVTDVRFIDNMPLQEGIDISQNNLKDISFLSKKNLTIVAAADNGIEDIRSLGOLP

```

```

/gene="bira"
/note="synonym: Spym3_1036"
7244..8185
/gene="bira"
/note="best blastp hit: gb|AAK34189.1| (AE006574) putative
biotin operon repressor [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product-"putative biotin operon repressor"
/protein_id-"AAM79643.1"
/db_xref="GI:21904773"
/translation-"MKTSEKITYQLISQTDDEPVSGEYVLADQUSISFTSVKSIKLENO
GIQIDSLKHKGRVMOGDILPKTISOGIGMPVYTPHPSQSTOLDPAKQIEAHNSAPR
LYIAPSOBAKGRDLROFFSASTGCIYMSMYLKPVPYPADMPYTMVAASIVKISR
LTGIDTEIKWVDIYLGNNHKAAGILTEAITSVEGLTIDVTIGVGLNFVDEPEAIA
OKAGSLFTEKEPITRNDLIIDIKRFLSIPYVDHYKVKYKESIVLANKQVTFIENSQEK
RAIATDITDQGLIYQFENGDIQTRSEISLSMEN"
complement(8160..8444)
/gene="Spym3_1037"
complement(8160..8444)
/gene="Spym3_1037"
/note="best non-GAS blastp hit: gb|AAK74993.1| (AE007392)
hypothetical protein [Streptococcus pneumoniae TIGR4], and
gb|AAK9574.1| (AE008452) hypothetical protein
[Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product-"conserved hypothetical protein"
/protein_id-"AAM79644.1"
/db_xref="GI:21904774"
/translation-"MGKVLPLETFYLVFFHSTVEFCYKAMTIRQFLPMAFVCAFEF
YFNDLLSGNVLFLFWGLLFRDLRRVHTINQTLTKTILKANSPPKKKD"
complement(8448..10118)
/gene="dnaX"
/note="synonym: Spym3_1038"
complement(8448..10118)
/gene="dnaX"
/note="best blastp hit: gb|AAL97971.1| (AE010058) DNA
polymerase III gamma/tau subunits [Streptococcus pyogenes
MGAS8232]"
/codon_start=1
/transl_table=11
/product-"DNA polymerase III gamma/tau subunits"
/protein_id-"AAM79645.1"
/db_xref="GI:21904775"
/translation-"MYQALYRKYSQTFDEWVGQSVISTTLKQAVESGKISHAYLFSG
PRGTSKTAAKIFAKAMCNPVODPCNOCIDITNGSLSEVIEIDAASNNQVDE
IRDIRDKSTVAPSRATYRVYIIDEVHMLSTGAFNVLKLEPPTENNVITATTEFLK
IPATILSRVOREPEAKOKAIRBLAVYIDEGIAVYVDALNLTARRAREGMRBALS
IILQALSLSPDNQVITAIABEITGISMLADGVRYVSODATDALALETTIDSGK
MSRPRATDLITYLRDLVKAAGDNQASVAFDNLSTSIDRIEQTIVTVSHLSEIK
KSHPRIAEAMWTIOAQEQIOLSQANISGELISIEITELKELAKOOLQOLOSRPD
SLARSKDTPKRTASRYRVDVITLIKMEETVRSOOSROYLALAKNANMEIINDISAO
D RALMGSEPVANSENATILAFEAANAOVSNRNLMFMFGIMSKAGFSPNLAIVP
RTDRQHKRKEFAQCKSKSDSVQEQEVALDIPEDFLDLKINTIDP"
complement(10118..10681)
/gene="Spym3_1039"
complement(10118..10681)
/gene="Spym3_1039"
/note="best non-GAS blastp hit: gb|AAK06294.1| (AE006448)
hypothetical protein [Lactococcus lactis subsp.
lactis]"
/codon_start=1
/transl_table=11
/product-"conserved hypothetical protein"
/protein_id-"AAM79646.1"
/db_xref="GI:21904776"

```

/translation="WTEAKGSRKYSNRRRAIIKGNKSKRIEQYOLMIAQAKELFA
NESNPLAMISNALSALINMLNPSPVAGFYLLIGPFGRRVSCYHRIKGVCG
ESASORFTIINDVQAHNYISCDAAASEIYVPMVKRGLIGVLDLSLVADYDEV
DOEYLEAFVDLEKTFEFTFNNFVKN"
10760..11569
/gene="spym3_1040"
10760..11569
/gene="spym3_1040"
/note="Best non-GAS blastp hit: gblAK98831.1 (AE008387)"

Search completed: September 4, 2003, 20:27:24
Job time : 3656 secs

Query Match 26.9%; Score 260; DB 1; Length 51454;
Best Local Similarity 56.8%; Pred. No. 2.7e-38;
Matches 504; Conservative 0; Mismatches 375; Indels 9; Gaps 1;
gene
cDS
34 AAAAAATTATTCAGAGTGCATCAGCTATATTCAGTACAGCTTACAGCTTGTTCG 93
31715 AATAACATCATGCTAGTGTGTGACATTTGGCTCAGTGTGCTTTACAGCTTGTCAA 31656
94 AAAGGTCAGAAAGTGCAGACCTTATCAGCATGAAAGGATGTCTTACAGACATCAA 153
31655 TCACATAATGACAAATACATAGTTATTTGATGAAAGGATGATACATTAAGCTTAGTAT 31596
154 TTTTATGACAGTAAAAAGCAACCTTAGCCCAACAAGTCTTGTAAATATGACATC 213
31595 TTTTACATGAAACAAACAAACAGAAAGTATCGCAAAAGCAATGCTAAATCTGTAAAT 31536
214 CAAAAGTTTGTGAATAACATATGAGCTCAGAGCTTATGATGAAGAGTGTATGATACT 273
31535 AGTCGTGTTTGAAGCTCAATATGTGTATTAAGGTTTCAAAAAAGAAAGTTGAAAAAGCG 31476
274 ATTGCCGAAGAAAAAACAAATATGCGGAAAACTAACAAAGTGTCTTGTACAGCAGGT 333
31475 TATCATTAACAAAGCTGAAACAGATATGCGCTTCTGCTGCTTGTGACACATCAAGC 31416
334 ATGACTCTGAAACAGCTAACTCAATTCGTACAGTAATTAATTAAGTTGACATTTG 393
31415 TTGACACCTGAGAGCTTTAAGCGCTAGATCCGCTCTTCAAAATTAAGTAATATGCGGTT 31356
394 AAGAAGGTAGCAGAGCTGAATTTGACAGATGAAGCTTATGAAGAGCTTTGATGATAC 453
31355 AAGAAGCAGCTAATAAAGAAATTTGACACACAAAGATATTAAGAAAGCATATGATCTTAT 31296
454 ACTCAGATGTAAAGGCTCAAAATCATCGCTTAAATATGAAGATTAAGGCCAAGAAATT 513
31295 ACTCCAACATATGCGAGCTCAAAATGATTACTTAAATGAAGAGACCTAAATCAGTC 31236
514 CTCGAAAAAGCCAGGACAGAGTGTGATTTTGTCAATTAAGCCAAAGATTAATCACT 573
31235 TTAGAGGAACCTAAAGCCGAGCGCACTTTACAGCTATGCTAAAGAAAAAACACAC- 31177
574 GATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGAATTCGCTTCAACAGAAATTA 633
31176 -----AACACCTGAGAAAAAAGTACCTTAATTAATTTGATTCAAGCGCGACAAATGTA 31125
634 CCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTGAATGTGATGTTCTGATGTGAT 693
31124 CCGACTGATGTGTAAGCGGCTTCAAGTTGAATGAGGTGCGATATCAGACGTTATC 31065
694 ACAGCAACTGCGACACACCTTACAGTACCAATATTAATGTAATAACTCATAAGAAA 753
31064 TCGGTTTGAATCCAACTTCTTAACAAAGATTTTACATGTTAAGGTGACTAAAAAA 31005
754 ACAGAAATATCATATATATGATGATCAAAAGAAATTAATAAAGCTTATCTGACT 813
31004 GCAGAAAAAATATCATGATGCGCAAGATATTAAGAAAGCTTTGAAGCTATCATTTAGCT 30945
814 CAAAAAATATGATTCACATTTGTTCAAGCATTAAGCAATTAAGCAAGTTCAGCAGCC 873
30944 GAAAAATCAAAAGATATGATTTCAAAACCAAGTTATTCGAATGATGATTAAGCT 30885
874 AATATCAAGTTAAGCAAGCTTCCAAATATCTTTACCAATAT 921
30884 AATGTAATAATTAAGCAAAAGCTTTGCTAATATTTTGGCGCAATAT 30837

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

pneumoniae useful for treating or preventing microbial infections - specifically streptococcal infections -

Disclosure; Fig 1; 18pp; English.

Streptococcus pneumoniae is a major cause of invasive diseases such as meningitis, bacteraemia and pneumonia, as well as non-invasive diseases such as acute otitis media and sinusitis. The present invention relates to a vaccine comprising Protease Maturation Protein (Pmp) of *S. pneumoniae*. The vaccine can be used for the treatment of *S. pneumoniae* infections. The present sequence is used for the sequence of *S. pneumoniae* pmp, which can be used to produce the coding sequence of the present invention. pmp is a surface-associated protein.

SQ Sequence 966 BP; 374 A; 170 C; 187 G; 235 T; 0 other;

Sequence 966 BP; 374 A; 170 C; 187 G; 235 T; 0 other;

Query Match	100.0%	Score 966;	DB 22;	Length 966;
Best Local Similarly	100.0%;	Pred. No. 4.2e-213;		
Matches 966; Conservative	0;	Mismatches	0;	Gaps 0;

OY	1	AGTAACACTTATATCTAAAGGAGTACACATGAGAAAAAATTTATTTGGCAGTGGCATACACA	60
Db	1	AGTAACACTTATATCTAAAGGAGTACACATGAGAAAAAATTTATTTGGCAGTGGCATACACA	60
OY	61	CTATTATCAGTAGCAACCTTTAGCAGCTTGTGTGAAAGGGGTGAGAAAGTGCAGACCTTATC	120
Db	61	CTATTATCAGTAGCAACCTTTAGCAGCTTGTGTGAAAGGGGTGAGAAAGTGCAGACCTTATC	120
OY	121	AGCATGAAAGGGGAGTGCATTTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTT	180
Db	121	AGCATGAAAGGGGAGTGCATTTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTT	180
OY	181	TCAGCCCAACAAGCTCTGTGTTAAATATGACACATCCAAAAAGTTTGTGAAAAACATAATGGC	240
Db	181	TCAGCCCAACAAGCTCTGTGTTAAATATGACACATCCAAAAAGTTTGTGAAAAACATAATGGC	240
OY	241	TCAGAGCTGTATGATTAAGAGAGTGTATGATTAATTTGCCGAGAAAAAACAATATGGC	300
Db	241	TCAGAGCTGTATGATTAAGAGAGTGTATGATTAATTTGCCGAGAAAAAACAATATGGC	300
OY	301	GAAAACTACCAACGCTGTCTGCACAGAGAGTATGACTTGAAMAACAGTAAGCTCA	360
Db	301	GAAAACTACCAACGCTGTCTGCACAGAGAGTATGACTTGAAMAACAGTAAGCTCA	360
OY	361	ATTGCTACAACTAAATTAATAGTTAGTTGGCAGTTAAGAAAGTAGCAGAACTGTAATTGACA	420
Db	361	ATTGCTACAACTAAATTAATAGTTAGTTGGCAGTTAAGAAAGTAGCAGAACTGTAATTGACA	420
OY	421	GATGACAGCCTATAGAAAGCCCTTTATGATGATCACTCCAGATGTAAACGGCTCAATATC	480
Db	421	GATGACAGCCTATAGAAAGCCCTTTATGATGATCACTCCAGATGTAAACGGCTCAATATC	480
OY	481	CGTCTTAATTAATGAAGATTAAGGCCAAGAAGTTCTCGAAAAAGCCCAAGCAGAGGTGCT	540
Db	481	CGTCTTAATTAATGAAGATTAAGGCCAAGAAGTTCTCGAAAAAGCCCAAGCAGAGGTGCT	540
OY	541	GATTTTGCCTAATTAGCCCAAGATTAATTCACCTGATGAAAAAACAAMAGAAATGCTGGA	600
Db	541	GATTTTGCCTAATTAGCCCAAGATTAATTCACCTGATGAAAAAACAAMAGAAATGCTGGA	600
OY	601	GAAATTAACCTTGTGATTTGCTTCAACAGAAATGACTGAGCAAGTCAAAAAAGCCGCTTTC	660
Db	601	GAAATTAACCTTGTGATTTGCTTCAACAGAAATGACTGAGCAAGTCAAAAAAGCCGCTTTC	660
OY	661	GCTTTAGATGAGATGATGTTTCTGATGATTACAGCAACGGGACACAGAGCTACAGT	720
Db	661	GCTTTAGATGAGATGATGTTTCTGATGATTACAGCAACGGGACACAGAGCTACAGT	720
OY	721	AGCCAAATTAATCAATTTGAAAACTCACTAAGAAACAGAAAAATCAATTAATTTGATGAC	780
Db	721	AGCCAAATTAATCAATTTGAAAACTCACTAAGAAACAGAAAAATCAATTAATTTGATGAC	780
OY	781	TACAAAGAAAAATTAATAAGTGTATCTTGACTCAAAAAACAATGATTAACAATTTGTT	840

Db	781	TACAAGAAAAATTTAAAACTGTATTCTTGACTCAAAAAACAATAATATTCACATTTGTT	840
QY	841	CAAAACATTTATCGSAAAAAATTTGCAAGACGCAATATCAAGTTAAAGCACCAAGCCTTC	900
Db	841	CAAAACATTTATCGSAAAAAATTTGCAAGACGCAATATCAAGTTAAAGCACCAAGCCTTC	900
QY	901	CAAAATATCTTTACCCCAATATATCGTGCGTGGAGATTCAAGCTCAAGCAGTAGTACATCA	960
Db	901	CAAAATATCTTTACCCCAATATATCGTGCGTGGAGATTCAAGCTCAAGCAGTAGTACATCA	960
QY	961	AACGAA 966	
Db	961	AACGAA 966	

RESULT 2	
AAV52241	
ID	AAV52241 standard; DNA; 11309 BP.

AC	AAV52241;
XX	
DT	23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:108.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds

Streptococcus pneumoniae

PN W09818931-A2

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19588.

PR 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

WPI: 1998-272225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae

PS Claim 1; Page 794-800; 1409pp; English

CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment of a sequence at least 95%
CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC *Streptococcus pneumoniae*. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the *S. pneumoniae* genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the *S. pneumoniae* genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the *S. pneumoniae* genome of commercial importance, or
CC expression modulating fragments of the *S. pneumoniae* genome. Products

CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for *S. pneumoniae*.

xx Sequence 11309 BP; 3557 A; 2073 C; 2451 G; 3228 T; 0 other;

Query Match 100.0%; Score 966; DB 19; Length 11309;
 Best Local Similarity 100.0%; Pred. No. 8,2e-213;
 Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ACTAACACTTATCTCAAGAGTAGACATGAAGAAAAATTTATGTCAGTGGCATACACA 60
DB 7632 AGTAACACTTATCTCAAGAGTAGACATGAAGAAAAATTTATGTCAGTGGCATACACA 7691
OY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTCAGAAAGTCAGACCTTATC 120
DB 7692 CTATTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTCAGAAAGTCAGACCTTATC 7751
OY 121 ACATGAAAAGGGGATGTCATTACAGACATCAATTTTATGACCACTGAAAAGCAACCT 180
DB 7752 ACATGAAAAGGGGATGTCATTACAGACATCAATTTTATGACCACTGAAAAGCAACCT 7811
OY 181 TCAGCCCAACAGTCTGTTAAATATGACATCCAAAAAGTTTGTGAAAAACATATGGC 240
DB 7812 TCAGCCCAACAGTCTGTTAAATATGACATCCAAAAAGTTTGTGAAAAACATATGGC 7871
OY 241 TCAGAGCTTGATTAAGAGGTTGATGATCTATTGCGAAGAAAAAACAAATATGGC 300
DB 7872 TCAGAGCTTGATTAAGAGGTTGATGATCTATTGCGAAGAAAAAACAAATATGGC 7931
OY 301 GAAAACCTACCAACGTCCTTCTCACAAGCAGTATGCTCTGAAAACAGCTAAAGCTCAA 360
DB 7932 GAAAACCTACCAACGTCCTTCTCACAAGCAGTATGCTCTGAAAACAGCTAAAGCTCAA 7991
OY 361 ATTGCTACAGTAATTAATGAGTGGCAGTTAAAGAGTACAGCAAGCTGAATTGACA 420
DB 7992 ATTGCTACAGTAATTAATGAGTGGCAGTTAAAGAGTACAGCAAGCTGAATTGACA 8051
OY 421 GATGAAGCCTATTAAGAAAGCCTTTGATGATGACACTCCAGATGTAACGGCTCAATATC 480
DB 8052 GATGAAGCCTATTAAGAAAGCCTTTGATGATGACACTCCAGATGTAACGGCTCAATATC 8111
OY 481 GCTCTTAATTAATGAAGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAGCAAGGTGCT 540
DB 8112 GCTCTTAATTAATGAAGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAGCAAGGTGCT 8171
OY 541 GATTTTGTCTCAATTAAGCCAAAGATATTCAGCTGATGAAAAAACAAGAAATGCTGGA 600
DB 8172 GATTTTGTCTCAATTAAGCCAAAGATATTCAGCTGATGAAAAAACAAGAAATGCTGGA 8231
OY 601 GAAATTAACCTTGTGATTCCTTCACAGAGTACCTGACCAAGTCAAAAAAGCCGCTTTC 660
DB 8232 GAAATTAACCTTGTGATTCCTTCACAGAGTACCTGACCAAGTCAAAAAAGCCGCTTTC 8291
OY 661 GCTTTAGATGATGAGTGTCTTCTGATGATTAAGCAACATGCGCACAGAGCTTCACT 720
DB 8292 GCTTTAGATGATGAGTGTCTTCTGATGATTAAGCAACATGCGCACAGAGCTTCACT 8351
OY 721 AGCCATATTAATTAATTAAGCTCACTAAGAAAAAGAAAAATCATCTATATGATGAC 780
DB 8352 AGCCATATTAATTAATTAAGCTCACTAAGAAAAAGAAAAATCATCTATATGATGAC 8411
OY 781 TACAAAAGAAAAATTAAGAACTGTTATCTTACTCAAAAAACAAATGATTCACATTTGTT 840
DB 8412 TACAAAAGAAAAATTAAGAACTGTTATCTTACTCAAAAAACAAATGATTCACATTTGTT 8471
OY 841 CAAAGATATTAAGGAAAAATTAAGCAAGCAATATCAAGGTTAAGCAAGCAAGCTTTC 900
DB 8472 CAAAGATATTAAGGAAAAATTAAGCAAGCAATATCAAGGTTAAGCAAGCAAGCTTTC 900
OY 901 CAAATATCTTTATACCAATATATCGGTGATGAGATTAAGCTCAAGCAAGTATGATCA 960
DB 8532 CAAATATCTTTATACCAATATATCGGTGATGAGATTAAGCTCAAGCAAGTATGATCA 8591

```

OY 961 AACGAA 966
 DB 8592 AACGAA 8597

RESULT 3*
 ID ABS56454 standard; DNA; 2162598 BP.
 XX ABS56454;

DT 10-FEB-2003 (first entry)

XX Streptococcus pneumoniae type 4 strain complete genome.

DE ds: bacterial meningitis; pneumonia; sepsis; otitis media; genome;
 KW ear infection; antilinfammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

XX MO200277021-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WC-1B02163.

XX 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelin H, Fraser C;

XX MPI; 2003-040579/03.

PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to *Streptococcus bacteria*, such as pneumonia, sepsis, otitis media
 PT or ear infection

PS Claim 17; SEQ ID No 4979; 56bp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a *Streptococcus nucleic acid* sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified. assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a *Streptococcus pneumoniae bacterium*, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC *Streptococcus bacteria*, particularly *S. pneumoniae*, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is the
 CC *Streptococcus pneumoniae* type 4 strain genome sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

```

XX Sequence 2162598 BP; 654373 A; 427176 C; 431369 G; 649680 T; 0 other;
SQ
Query Match      100.0%; Score 966; DB 25; Length 2162598;
Best Local Similarity 100.0%; Pred. No. 3.4e-212;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGTGCATCACA 60
DB 820773 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGTGCATCACA 820832
QY 61 CTTATTCAGTAGCAACTTATAGCAGCTTGTGAAAAGGTCAGAAAGTGCAGACCTTATC 120
DB 820833 CTTATTCAGTAGCAACTTATAGCAGCTTGTGAAAAGGTCAGAAAGTGCAGACCTTATC 820892
QY 121 ACGATGAAGGGGATGCTCATTAAGAAACATCAATTTATGAGCAAGTGAAGAACACCT 180
DB 820893 ACGATGAAGGGGATGCTCATTAAGAAACATCAATTTATGAGCAAGTGAAGAACACCT 820952
QY 181 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGTAAAAACAATATGCG 240
DB 820953 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGTAAAAACAATATGCG 821012
QY 241 TCAGAGCTTGATGATTAAGAGTGTGATGATGATGATGATGATGATGATGATGATG 300
DB 821013 TCAGAGCTTGATGATTAAGAGTGTGATGATGATGATGATGATGATGATGATGATG 821072
QY 301 GAAAGCTACCAAGCTGTCTGTGCACAGAGATGATGATGATGATGATGATGATGATGATG 360
DB 821073 GAAAGCTACCAAGCTGTCTGTGCACAGAGATGATGATGATGATGATGATGATGATGATG 821132
QY 361 ATTCCGACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
DB 821133 ATTCCGACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 821192
QY 421 GATGAAGCTTATTAAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 480
DB 821193 GATGAAGCTTATTAAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 821252
QY 481 CGTCTTAATTAATTAAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 540
DB 821253 CGTCTTAATTAATTAAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 821312
QY 541 GATTTGGCTCAATTAACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 600
DB 821313 GATTTGGCTCAATTAACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 821372
QY 601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
DB 821373 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 821432
QY 661 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
DB 821433 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 821492
QY 721 AGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
DB 821493 AGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 821552
QY 781 TACAAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
DB 821553 TACAAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 821612
QY 841 CAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
DB 821613 CAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 821672
QY 901 CAAAAATATCTTATCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
DB 821673 CAAAAATATCTTATCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 821732
QY 961 AACGAA 966
DB 821733 AACGAA 966

```

```

DB 821733 AACGAA 821738
RESULT 4
ABX06679
ID ABX06679 standard; DNA; 939 BP.
XX
XX ABX06679;
XX
XX 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #967.
XX
XX Gene; ds: bacterial meningitis; pneumonia; sepsis; otitis media;
XX ear infection; antinflammatory; antibacterial; immunostimulant;
XX auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae type 4 strain.
XX
XX MO20027021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002MO-IB02163.
XX
XX 27-MAR-2001; 2001GB-0007658.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masinani V, Tetteijn H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX
XX P-PSDB; AB001392.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
XX or ear infection
XX
XX Claim 6; SEQ ID NO 1933; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX ABS56454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2489 identified coding region from the genomic sequence.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX

```

Sequence 939 BP; 363 A; 165 C; 182 G; 229 T; 0 other;

Query Match 97.2%; Score 939; DB 25; Length 939;
Best Local Similarity 100.0%; Pred. No. 6.9e-207;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 28 ATGAGAGAAAAATTTTGGCAGAGTCCCATCATCTATTATTCAGTACCACTTTAGCAGCT
DB 1 ATGAGAGAAAAATTTTGGCAGAGTCCCATCATCTATTATTCAGTACCACTTTAGCAGCT 60
OY 88 TGTTCGAAAGGGTGCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 147
DB 61 TGTTCGAAAGGGTGCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 120
OY 148 CATCAATTTTATGACAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATG 207
DB 121 CATCAATTTTATGACAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATG 180
OY 208 ACCATCCAAAAAGTTTTTGAAGAAATATGCTCAGAGCTTGATGATTAAGAGTTGAT 267
DB 181 ACCATCCAAAAAGTTTTTGAAGAAATATGCTCAGAGCTTGATGATTAAGAGTTGAT 240
OY 268 GATCTATTGTCGCGAAGAAAAAACAATATGCGCAAACTACCAAGCTCTTGTACAA 327
DB 241 GATCTATTGTCGCGAAGAAAAAACAATATGCGCAAACTACCAAGCTCTTGTACAA 300
OY 328 GCAGGTATGACTCTTGAAGAACGTAAGCTCAATTCGACAGTAAGTAATTTAGTTGAGTTG 387
DB 301 GCAGGTATGACTCTTGAAGAACGTAAGCTCAATTCGACAGTAAGTAATTTAGTTGAGTTG 360
OY 388 GCAGTTAAGAAAGGTAGCAGAAAGCTGATGACAGATGAAAGCTATTAAGAAAGCTTTGAT 447
DB 361 GCAGTTAAGAAAGGTAGCAGAAAGCTGATGACAGATGAAAGCTATTAAGAAAGCTTTGAT 420
OY 448 GAGTACACTCCGAGATGTAAGGCTCAAAATCATCCGCTTATTAATGAAGATTAAGGCCAA 507
DB 421 GAGTACACTCCGAGATGTAAGGCTCAAAATCATCCGCTTATTAATGAAGATTAAGGCCAA 480
OY 508 GAGTCTTCGAAAGAACCCAGCAGAGAGGTGCTGATTTGCTCAATTTAGCCAAAGATAT 567
DB 481 GAGTCTTCGAAAGAACCCAGCAGAGAGGTGCTGATTTGCTCAATTTAGCCAAAGATAT 540
OY 568 TCACTGATGAAAAAACAAGAAATGCTGAGAAATATACCTTTGATTCGCTTCACAA 627
DB 541 TCACTGATGAAAAAACAAGAAATGCTGAGAAATATACCTTTGATTCGCTTCACAA 600
OY 628 GAGTACCTGAGCAAGTCAAAAAGCCGCTTCGCTTGAATGATGATGCTGTTTGAT 687
DB 601 GAGTACCTGAGCAAGTCAAAAAGCCGCTTCGCTTGAATGATGATGCTGTTTGAT 660
OY 688 GTGATTACAGCAACTGGCACAGACCTTACAGTACGCAATTTTAAACCTCAGCT 747
DB 661 GTGATTACAGCAACTGGCACAGACCTTACAGTACGCAATTTTAAACCTCAGCT 720
OY 748 AAGAAAACAGAAAAATCATCTAATATGATGACTACAAAGAAAAATTTAAAACTGTATC 807
DB 721 AAGAAAACAGAAAAATCATCTAATATGATGACTACAAAGAAAAATTTAAAACTGTATC 780
OY 808 TTGACTCAAAAACAAATGATTCACATTTGTTCAAGCATATTCGAAAGAAATTCGAA 867
DB 781 TTGACTCAAAAACAAATGATTCACATTTGTTCAAGCATATTCGAAAGAAATTCGAA 840
OY 868 GCAGCAATATATCAAGTTAAGACCAAGCTTCGAAATATCTTTACCAATATATCGGT 927
DB 841 GCAGCAATATATCAAGTTAAGACCAAGCTTCGAAATATCTTTACCAATATATCGGT 900
OY 928 GGTGAGATTCAGCTCAAGCAGTACATCAAAAGAA 966
DB 901 GGTGAGATTCAGCTCAAGCAGTACATCAAAAGAA 939

```

RESULT 5
AAA05430

```

ID AAA05430 standard; DNA; 942 BP.
XX
AC AAA05430;
XX
XX 24-MAY-2000 (first entry)
DT
XX Streptococcus pneumoniae nucleotide sequence ID212 - 4136.1.
DE
XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease; ds.
XX Streptococcus pneumoniae.
OS
XX MO200006737-A2.
PN
XX 10-FEB-2000.
PD
XX
XX 27-JUL-1999; 99MO-GB02451.
PF
XX 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX
XX (MICR-) MICROBIAL TECHNIQS LTD.
PA
XX Gilbert CFG, Hansbro PM;
PI
XX WPI: 2000-195300/17.
DR P-PSDB; AY81675.
XX
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX
XX Claim 6; Page 92; 108pp; English.
PS
XX
XX AA81501 to AA81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AA05591 to AA05614 represent primers used in the
CC exemplification of the present invention.
XX
SQ Sequence 942 BP; 364 A; 165 C; 183 G; 230 T; 0 other;

```

Query Match 97.2%; Score 939; DB 21; Length 942;
Best Local Similarity 100.0%; Pred. No. 6.9e-207;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 28 ATGAGAGAAAAATTTTGGCAGAGTCCCATCATCTATTATTCAGTACCACTTTAGCAGCT 87
DB 1 ATGAGAGAAAAATTTTGGCAGAGTCCCATCATCTATTATTCAGTACCACTTTAGCAGCT 60
OY 88 TGTTCGAAAGGGTGCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 147
DB 61 TGTTCGAAAGGGTGCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 120
OY 148 CATCAATTTTATGACAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATG 207
DB 121 CATCAATTTTATGACAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATG 180
OY 208 ACCATCCAAAAAGTTTTTGAAGAAATATGCTCAGAGCTTGATGATTAAGAGTTGAT 267
DB 181 ACCATCCAAAAAGTTTTTGAAGAAATATGCTCAGAGCTTGATGATTAAGAGTTGAT 240
OY 268 GATCTATTGTCGCGAAGAAAAAACAATATGCGCAAACTACCAAGCTCTTGTACAA 327

```

```

Db      241 GATGCTATTTGGCGAAGAAAAAACAATATGCGCAAACTACCAAGCTGCTGTCACAA 300
Oy      328 GCAGGTATGACTCTTGAAGACGCTTAAGCTCAATTCCTACAGTAATTTAGTTAGTTG 387
Db      301 GCAGGTATGACTCTTGAAGACGCTTAAGCTCAATTCCTACAGTAATTTAGTTAGTTG 360
Oy      388 GCAGTTAAGAAAGTGAAGACGCTTAATTCAGATGAAGCGCTTAAGAAAGCGCTTTGAT 447
Db      361 GCAGTTAAGAAAGTGAAGACGCTTAATTCAGATGAAGCGCTTAAGAAAGCGCTTTGAT 420
Oy      448 GAGTACTCTCCAGATGTAAGCGCTCAAAATCATCGCTTTAATATGAAGATTAAGGCCAAA 507
Db      421 GAGTACTCTCCAGATGTAAGCGCTCAAAATCATCGCTTTAATATGAAGATTAAGGCCAAA 480
Oy      508 GAACTTTCGAAAAAGCCAGACAGAGGCTGCTGATTTTTCCTCAATTTGCGCAAGATATAT 567
Db      481 GAACTTTCGAAAAAGCCAGACAGAGGCTGCTGATTTTTCCTCAATTTGCGCAAGATATAT 540
Oy      568 TCACTGATGAAAAAACAAGAAATGGTGAGAAATTAACCTTTGATTCTGCTTCAACA 627
Db      541 TCACTGATGAAAAAACAAGAAATGGTGAGAAATTAACCTTTGATTCTGCTTCAACA 600
Oy      628 GAACTGCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGATGATGATGATGATGAT 687
Db      601 GAACTGCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGATGATGATGATGATGAT 660
Oy      688 GGAATTAAGCAAGTGGCAGACAGAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 747
Db      661 GGAATTAAGCAAGTGGCAGACAGAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 720
Oy      748 AAGAAAAAGAAAAATCATCTAATATGATGATGATGATGATGATGATGATGATGATGAT 807
Db      721 AAGAAAAAGAAAAATCATCTAATATGATGATGATGATGATGATGATGATGATGATGAT 780
Oy      808 TTGACTCAAAAAACAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 867
Db      781 TTGACTCAAAAAACAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Oy      868 GGAAGCAATATCAAGTGAAGACCAAGGCTTCAAAATATGATGATGATGATGATGATGAT 927
Db      841 GGAAGCAATATCAAGTGAAGACCAAGGCTTCAAAATATGATGATGATGATGATGATGAT 900
Oy      928 GGTGAGATTCAGCTCAAGCAAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 966
Db      901 GGTGAGATTCAGCTCAAGCAAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 939

```

RESULT 6

AAV27340 standard; DNA; 640 BP.

AAV27340;

02-OCT-1998 (first entry)

Streptococcus pneumoniae SP0021 nucleotide.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KM detection; pneumonia; otitis media; meningitis; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers

FT CDS

FT /tag- a

FT /product= "SP0021"

FT /note= "no stop codon given"

XX

XX

XX

```

PF      30-OCT-1997; 97WO-US19422.
XX      31-OCT-1996; 96US-0029960.
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX      WPI: 1998-27224/24.
DR      P-PSDB: AAM55079.
XX      Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT      pneumoniae - or their epitope-containing fragments, useful in
PT      protective or therapeutic vaccines, and for diagnosis
XX      Claim 1; Page 55; 118pp; English.
CC      The present sequence encodes a protein from Streptococcus pneumoniae.
CC      The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC      can be useful in vaccines for inducing protective antibodies against
CC      Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC      pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC      are used to detect Streptococcus infection (by usual hybridisation or
CC      amplification methods), also for isolating Streptococcus genes or their
CC      allelic variants. The protein can be used similarly to detect specific
CC      antibodies in standard immunoassays, especially for diagnosing or
CC      monitoring infections. Antibodies which bind the protein are used to
CC      detect corresponding antigens, to purify the protein and for passive
CC      immunisation (optionally coupled to a toxin). Vaccines are administered,
CC      e.g. by injection, orally or through the skin, typically at 0.01-1000
CC      (especially 10-300) mu g/ml per dose.
XX      Sequence 640 BP; 234 A; 112 C; 141 G; 153 T; 0 other;
SQ
Query Match      57.2%; Score 552.4; DB 19; Length 640;
Best Local Similarity 94.7%; Pred. No. 8.2e-118;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;
Oy      90 TTGAAAGGGTCAGAAAGTGCAGACCTTATCGATGAAAGGGATGCTATTACGAACA 149
Db      1 TTGAAAGGGTCAGAAAGTGCAGACCTTATCGATGAAAGGGATGCTATTACGAACA 60
Oy      150 TCAATTTTATGAGCAAGTGAAGCAAGCAAGCTTCAAGCAAGCTTGAATATATGAC 209
Db      61 TCAATTTTATGAGCAAGTGAAGCAAGCAAGCTTCAAGCAAGCTTGAATATATGAC 120
Oy      210 CATCCAAAAAGTTTGAAGAAACAATATGCTCAGAGCTTGATGATTAAGAGCTTGATGA 269
Db      121 CATCCAAAAAGTTTGAAGAAACAATATGCTCAGAGCTTGATGATTAAGAGCTTGATGA 180
Oy      270 TACTATGCGCGAAGAAAAAACAATATGCGGAAAAAATACCAAGCTGCTTGTCACAAGC 329
Db      181 TACTATGCGCGAAGAAAAAACAATATGCGGAAAAAATACCAAGCTGCTTGTCTCAAGC 240
Oy      330 AGGTATGACTCTGGAACACGTAAGGCTCAAAATTCGTACAACTAATTAATTAAGTGAAGTGGC 389
Db      241 AGGTATGACTCTGGAACACGTAAGGCTCAAAATTCGTACAACTAATTAATTAAGTGAAGTGGC 300
Oy      390 AGTTAAGAAAGTGAAGCAAGTGAATTCACATGAAGCTTTAAGAAAGCTTTGATGA 449
Db      301 AGTTAAGAAAGTGAAGCAAGTGAATTCACATGAAGCTTTAAGAAAGCTTTGATGA 360
Oy      450 GTACACTCCAGATGTAAGGCTCAATCATCGCTTATTAATGAAGATTAAGGCCAAGA 509
Db      361 GTACACTCCAGATGTAAGGCTCAATCATCGCTTATTAATGAAGATTAAGGCCAAGA 420
Oy      510 AGTTTCGAAAAAGCCAGACGAGAGTCTGATTTTGTCTCAATTAGCCAAAGATTAATTC 569
Db      421 AGTTTCGAAAAAGCCAGACGAGAGTCTGATTTTGTCTCAATTAGCCAAAGATTAATTC 480
Oy      570 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTTGATTCGCTTCAACAGA 629
Db      481 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTTGATTCGCTTCAACAGA 540

```


QY 630 AGTACCTG--AGCAGTCACAAAAGCCGCTTTCGCTTAGATGTGATGTCTTCTGAT 687
 DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTAGATGTGATGTCTTCT 600
 QY 688 G-----TGATTACAGCAACTGGCACACA 710
 DB 601 GGATGTGGATTACAGCAACTGGGCACACA 628

RESULT 7

AB084808
 ID AB084808 standard; DNA; 640 BP.

AB084808;

04-SEP-2002 (first entry)

S. pneumoniae SP021 nucleotide sequence SEQ ID NO:33.

Streptococcus pneumoniae: epitope; vaccine; antigenic protein;
 antibacterial; Streptococcal infection; detection; gene; ds.

Streptococcus pneumoniae.

US2002061545-A1.

23-MAY-2002.

22-JAN-2001; 2001US-0765272.

30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;

DR WPI: 2002-479261/51.
 DR P-PSDB; ABP54573.

PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -

PS Claim 1; Page 24; 70pp; English.

CC AB084792 to AB084904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.

XX Sequence 640 BP; 234 A; 112 C; 141 G; 153 T; 0 other;

Query Match 57.2%; Score 552.4; DB 24; Length 640;

Best Local Similarity 94.7%; Pred. No. 8.2e-118;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTGGAAGGTCAGAAAGTGCAGACTTATCGATGAAGGGGATGTCATTACAGACA 149
 DB 1 TTGGAAGGTCAGAAAGTGCAGACTTATCGATGAAGGGGATGTCATTACAGACA 60

QY 150 TCAATTTTATGAGCAAGTGAAGCAACCCCTTACGCCCAACAGTCTTGTAAATATGAC 209
 DB 61 TCAATTTTATGAGCAAGTGAAGCAACCCCTTACGCCCAACAGTCTTGTAAATATGAC 120
 QY 210 CATCCAAAAGTTTGTAAAAAACAATATGCTCAGAGCTTGATGTAAGAAGCTTGATGA 269
 DB 121 CATCCAAAAGTTTGTAAAAAACAATATGCTCAGAGCTTGATGTAAGAAGCTTGATGA 180
 QY 270 TACTATTTGCCGGAAGAAAAAACAATATGCGGCAAAATGACAACTGTCTGTCAAGC 329
 DB 181 TACTATTTGCCGGAAGAAAAAACAATATGCGGCAAAATGACAACTGTCTGTCAAGC 240
 QY 330 AGGTATGACTCTTGAAGACGTAAGCTCAATATGCTACAGTAATTAAGTGAAGTGGC 389
 DB 241 AGGTATGACTCTTGAAGACGTAAGCTCAATATGCTACAGTAATTAAGTGAAGTGGC 300
 QY 390 AGTTAAGAAAGTAGCAGAAAGCTGAATTGACAGATGAAGCTTAAGAAAGCCTTGTGATGA 449
 DB 301 AGTTAAGAAAGTAGCAGAAAGCTGAATTGACAGATGAAGCTTAAGAAAGCCTTGTGATGA 360
 QY 450 GTACACTCCAGATGTAAAGCGCTCAATCATCGCTTAATATGAAGTAAGCCCAAGA 509
 DB 361 GTACACTCCAGATGTAAAGCGCTCAATCATCGCTTAATATGAAGTAAGCCCAAGA 420
 QY 510 AGTTCTCGAAAAAGCCAAAGCAGAAAGTGTGATTTTCTCAATTAGCCAAAGATAATTC 569
 DB 421 AGTTCTCGAAAAAGCCAAAGCAGAAAGTGTGATTTTCTCAATTAGCCAAAGATAATTC 480
 QY 570 AACTGATGAAAAAAGCAAAAGAAATGTGTGAGAAATTAACCTTGTCTGCTCAACAGA 629
 DB 481 AACTGATGAAAAAAGCAAAAGAAATGTGTGAGAAATTAACCTTGTCTGCTCAACAGA 540
 QY 630 AGTACCTG--AGCAGTCACAAAAGCCGCTTTCGCTTAGATGTGATGTCTTCTGAT 687
 DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTAGATGTGATGTCTTCT 600
 QY 688 G-----TGATTACAGCAACTGGCACACA 710
 DB 601 GGATGTGGATTACAGCAACTGGGCACACA 628

RESULT 8

AA30925
 ID AA30925 standard; DNA; 451 BP.

AA30925;

20-MAY-1999 (first entry)

Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:202.

Streptococcus pneumoniae strain 0100993; vaccine; immune response;
 Streptococcal infection; pneumococcal; ss.

Streptococcus pneumoniae.

MO9737026-A1.

09-OCT-1997.

01-APR-1997; 97WO-US05306.

22-AUG-1996; 96US-0025788.

02-APR-1996; 96US-0014690.

(SMK) SMITHKLINE BEECHAM CORP.

(SMK) SMITHKLINE BEECHAM PLC.

Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

WPI: 1997-503111/46.
 P-PSDB; AAY11343.

XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT vaccines, drug screening, etc
XX

PS Claim 5: Page 172: 354pp: English.

XX AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC encode the novel proteins given in AAY1114 to AAY1137. The proteins,
CC isolated from Streptococcus pneumoniae, can be used in vaccines against
CC streptococcal infections and in assays for identifying compounds that
CC inhibit or activate the activity of the proteins. The antagonists can
CC be used to treat an individual having need to inhibit a bacterial
CC protein. Vectors expressing the proteins can be used to induce a
CC protective immune response in mammals.

XX Sequence 451 BP: 170 A; 77 C; 87 G; 116 T; 1 other;

Query Match 27.5%; Score 265.6; DB 19; Length 451;
Best Local Similarity 97.9%; Pred. No. 9.7e-52;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 1 ACTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTATGTCAGTGCCTATCACA 60
DB 167 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGTCAGTGCCTATCACA 226
OY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGACCTTATC 120
DB 227 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGACCTTATC 286
OY 121 AGCATGAAGAGGGAGTGTCTTACAGAACATTTATGACCAAGTGAAGAACACCT 180
DB 287 AGCATGAAGAGGGAGTGTCTTACAGAACATTTATGACCAAGTGAAGAACACCT 346
OY 181 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 239
DB 347 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGGAAAAACAATATG 406
OY 240 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATGCGCGAGA 284
DB 407 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATGCGCGAGA 451

RESULT 9

AAT98753
ID AAT98753 standard: DNA; 791 BP.

AC AAT98753;

DE 10-NOV-1998 (first entry)

XX DNA encoding a S. pneumoniae protein of unknown function.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KM pathogenesis; ss.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers

XX CDS 534..758

XX FT /tag= a

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-0507950.

XX 14-MAY-1996; 96US-0017670.

PA (SMK) SMITHKLINE BEECHAM CORP.
PA (SMK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;

XX WPI: 1998-008793/01.

DR P-PSDB; AAM38711.

PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections

PS Claim 4: Page 239; 483pp: English.

XX This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.

XX Sequence 791 BP: 258 A; 140 C; 168 G; 224 T; 1 other;

Query Match 27.5%; Score 265.6; DB 19; Length 791;
Best Local Similarity 97.9%; Pred. No. 1.1e-51;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 1 ACTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTATGTCAGTGCCTATCACA 60
DB 507 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGTCAGTGCCTATCACA 566
OY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGACCTTATC 120
DB 567 CTATTATCAGTAGCAACTTTAGCAGCTTGTGCAAGAGGTGCAAGACCTTATC 626
OY 121 AGCATGAAGAGGGAGTGTCTTACAGAACATTTATGACCAAGTGAAGAACACCT 180
DB 627 AGCATGAAGAGGGAGTGTCTTACAGAACATTTATGACCAAGTGAAGAACACCT 686
OY 181 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 239
DB 687 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGGAAAAACAATATG 746
OY 240 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATGCGCGAGA 284
DB 747 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATGCGCGAGA 791

RESULT 10

ABN69360
ID ABN69360 standard: DNA; 1053 BP.

AC ABN69360;

DE 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 6633.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

DR WPI: 2002-352536/38.
 DR P-PSDB: ABP30803.
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 7, Page 4184; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 927 BP; 370 A; 152 C; 161 G; 244 T; 0 other;
 XX
 Query Match 26.9%; Score 259.4; DB 24; Length 927;
 Best Local Similarity 57.5%; Pred. No. 3.2e-50;
 Matches 511; Conservative 0; Mismatches 366; Indels 12; Gaps 2;
 QY 36 AAAATTTATTTGGAGTGCCTTACACATTTATTCAGTACAGCACTTTCAGCTTGTTC-- 92
 DB 15 AAAATCTTGACGAGCTTTTGTGACATTTATTCAGTACAGCACTTTCAGCTTGTTCAGG 74
 QY 93 GAAAGGTCAGAGAGTGCAGCTTATTCAGATGAAAGGATGTCATTCAGCAACATCA 152
 DB 75 TAAACATCAATATGGAACAAATGTTTCAATGAAAGGCGACACTTATTCAGTCTGTA 134
 QY 153 ATTTATGAGCAAGTGAAGCAACCTTCACGCCCAACAGTCTTGTAAATATGACAT 212
 DB 135 TTTTATGATCAAGTAAACAAATCAAAAGCTGCACACAAATCAATGCTTACATTTGATCT 194
 QY 213 CCAAAAAGTTTGTGAAAAACATATGCTCAGACCTTGATGATTAAGAGGTTGATGATAC 272
 DB 195 CTCAGCTGTTTGTATACACAGTATGCTGATTAAGTTCGATTAAGAAAAAGTATCAGAGC 254
 QY 273 TATTGCCAGAAAAAATATATGCGGAAACATACCAACGTCGCTTGTCCAGACAG 332
 DB 255 TTATATAAGACAGCTAAAGCTATGCTATTCATTTTCAGCGCACTTTCAGACAGCAGG 314
 QY 333 TATGACTCTTGAACACAGCTAAAGCTCAATTCGTACAGTAATTAATTAGTTGAGTGCAGT 392
 DB 315 TTTGACTCCGGAAGGTTACAAACAAATTCGCACACTATGCTGTGGAATATGCTGT 374
 QY 393 TAAAGAGTAGAGAGAGCTGAATTCAGACATTAACCCRTTAAGAAAGCTTGTGATGAGTA 452
 DB 375 AAAAAGACAGCTAAGAAAGATTAACAGAACCAACATTAAGAAAGCTATTAAGAAACTA 434
 QY 453 CACTCCAGATGTAAAGGCTCAATATCCTCGCTTAATATAGAAAGTAAGGCCAAAGAGT 512
 DB 435 TACTCTGGAACCTTGTACAAAGTATCAATTAATGATGACAGAGTAAGTAAGTATCTGT 494
 QY 513 TCTCGAAAAAGCCAGAGAGAGTGTGATTTGCTCAATTAATGCAAAAGATTAATTAAC 572
 DB 495 CCTTAAGATGTAAAGGCTGATGAGGATTTTTCAGAAAGATTCAGAAAAAGAAAAACAAC 554
 QY 573 TGATAAAAACAAAGAAATATGTCGAGAAATTAACCTTGTGATTCCTTCAACAGAGT 632
 DB 555 AGCTACTGATTAAGAAAGT-----GAGTATTAATTTGATTCGACGAGGACAACCT 605

QY 633 ACCTGAGCAAGTCAAAAAAGCCGCTTGCCTTAGATGTGATGTGATGTGAT 692
 DB 606 CCTTAAGAAAGTATATTCAGACAGCTTTAAGCTAGTAAATGTGTTTCAGATGTG 665
 QY 693 TACAGCACTGCACAAAGCCTACAGTATGCCAATTTATCATTTGTAAGCACTAAGAA 752
 DB 666 TTCACAGGTTGATTCACCAACTATTAACAGTACTACATCATTAAGTAAGTATGATTA 725
 QY 753 AACAGAAATATCATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812
 DB 726 GACAGGAAAAATATGATTTGGAATCTTACAAATTCGCTTAAGAAAGTATTTGTA 785
 QY 813 TCAAAAACAAATGATTCACATTTGTTCAAAAGCATTATCGMAAAGAAATGCAACAGC 872
 DB 786 GGTAAACAGAGATAGAGCCCTTCAAAATTAAGATTTCAAAAGCCTTGAAGAAAGC 845
 QY 873 CATATCAAGTTAAGACCAAGCTTCCAAATATCTTACCAATAT 921
 DB 846 TATGTTAAGATTAAGATTAAGCAATTCAGGACATCTTATCAATAT 894

RESULT 12

ABN71519
 ID ABN71519 standard; DNA; 927 BP.

AC ABN71519;

XX 01-JUL-2002 (first entry)

DE Streptococcus polynucleotide SEQ ID NO 10951.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

KM anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.

OS Streptococcus agalactiae.

PN WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001MO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettein H;

XX WPI: 2002-352536/38.

XX P-PSDB: ABP30888.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 7, Page 4192; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

Sequence 927 BP; 370 A; 152 C; 161 G; 244 T; 0 other;

Query Match 26.9%; Score 259.4; DB 24; Length 927;
Best Local Similarity 57.5%; Pred. No. 3.2e-50;
Matches 511; Conservative 0; Mismatches 366; Indels 12; Gaps 2;

```

OY 36 AAAATTTATGGCAGTGCATCATATATATGAGTACAGTTCAGTCTGTC--- 92
DB 15 AAAACTTGCAGCAGGCTTTTACACTTATGTCAGTTCGACATTTAGCTGCTGCTCAGG 74
OY 93 GAAAGGTCAGAAAGTGCAGACCTTATCAGATGAAAGGGATGTCATTACGAACATCA 152
DB 75 TAAACATCAATGGAACAAATGTTGTCATGAAAGGCGACACTATACAGTCTCTGA 134
OY 153 ATTTATGAGCAAGTGAAGAACACCCCTTACGCCAACAAGCTTGTATTAATATGACAT 212
DB 135 TTTTATGATCAAGTGAAGAACATCAAAAGCTGCACACAAATCAATGCTTACATTTGATCCT 194
OY 213 CCAAAAGTTTTGAAGAAATATGCTCAGAGCTTGATGATTAAGAGTTCATGATATAC 272
DB 195 CTCAGGTGTTTGTATACACATATGTCATTAAGTTTCAGTAAAGATATCAGAAAC 254
OY 273 TATTCGCCGAAAGAAAAAATATGCGGAAATACCAAGCTGCTTGTCCAGAACAGG 332
DB 255 TTTATATAGACAGCTAAAGGCTATGTTATTCATTTTCAAGCGCACTTTCACAGCAGG 314
OY 333 TATGACTCTTGAACACGTAAGCTCAATTTGTCACATTAATATGATGATGTCAGT 392
DB 315 TTTGACTCCGGAAGTTACAAACAAATTTGCAACAACTATGCTTGTGAAATATGCTGT 374
OY 393 TAAAGAGTACGAGAGTGAATGACATGACAGCCCTTAAGAAAGCTTGTATGATGATGA 452
DB 375 AAAAGAGAGCTAAGAAAGATTAACAGAACAACTTAAGAAAGCATATATTAACATCA 434
OY 453 CACTCCAGATGTACGCTCAATATCATCCGTTTAATTAATGAAGATAGGCCAAGAGAT 512
DB 435 TACTCTGAACTTGTGTACAGTATCAATTTGGTACAGAGATTAAGCTTAATCTGT 494
OY 513 TCTCGAAAAGCGCAGGAGGCTGATTTTGTCTCAATAGCCAAAGATATATCAAC 572
DB 495 CCTTAAAGATGTAAGGCTGATGAGCTGATTTTGCAGAAAGATTGCAAAAAAGCAAC 554
OY 573 TGATGAAAAAACAAGAAATATGTCGAGAAATTAACCTTGTATGCTTCAACAGAGAT 632
DB 555 ACCTACTGATAGAAAGT-----GATATTAATTTGATTCGACGGGACAAACCT 605
OY 633 ACCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATGATGATGATGATGATGAT 692
DB 606 CCTTAAAGATGATGTCAGAGCCCTTAAAGCTAATAAAAGGCTTTCAGATGATGAT 665
OY 693 TACAGCACTGGCACAAGCTACAGTCAATATTTCAATTTTAACTCACTAAGAA 752
DB 666 TTCAACGGTTGATCAACAACTTATTAACAAAGTTACTATCAATTAAGTAATTA 725
OY 753 AACGAAAAATCATCTAATATTTGATGATGATGATGATGATGATGATGATGATGAT 812
DB 726 GACAGAGAAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
OY 813 TCAAAAAAATAATGATTCATATTTGTTCAAGCATTAATGGAAGAAATGCAAGCAGC 872
DB 786 GGAATAAACAGCAGATAGAGCTTCCAAATTAAGGATTTCAAAAGCCTTGAAGAAAGC 845
OY 873 CAATATCAGGTTAAGGACAGCCTTCCAAATATCTTACCAATAT 921
DB 846 TAAATGTTAAGATTAAGATTAAGCATTTGACAGCATCTTATACAAATAT 894

```

RESULT 13

ID ABN71527 standard; DNA; 2155561 BP.

AC ABN71527;

DT 02-JUL-2002 (first entry)

DE Streptococcus polynucleotide SEQ ID NO 10967.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

OS Streptococcus sp.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PI (GENO-) INST GENOMIC RES.

PI Tettein H;

XX WPI; 2002-352536/38.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 8; Page 4196-4488; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and anti-inflammatory

XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

XX antibodies that bind (I) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (I) are used to detect Streptococcus in a

XX biological sample. (I) is used to determine whether a compound binds to

XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

XX acid encoding (I) may be used to recombinantly produce (I) and may be

XX used in gene therapy. Antibodies to (I) are used for affinity

XX chromatography, immunoassays, and distinguishing/identifying

XX Streptococcus proteins.

XX Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 26.9%; Score 259.4; DB 24; Length 2155561;
Best Local Similarity 57.5%; Pred. No. 2.7e-49;
Matches 511; Conservative 0; Mismatches 366; Indels 12; Gaps 2;

```

OY 36 AAAATTTATGGCAGTGCATCATATATGAGTACAGTTCAGTCTGTC--- 92
DB 777887 GAAACTTGCAGCAGGCTTTTACACTTATGTCAGTTCGACATTTAGCTGCTGCTCAGG 777946
OY 93 GAAAGGTCAGAAAGTGCAGACCTTATCAGATGAAAGGGATGTCATTACGAACATCA 152
DB 777947 TAAACATCAATGGAACAAATGTTGTCATGAAAGGCGACACTATTAACAGTCTCTGA 778006

```

```

QY 153 ATTTATGAGCAAGTGAAGAAACCAACCCCTTCAGCCCAACAAGTCTGTTAATAATGACCAT 212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778007 TTTTATGATCAATGAAAAACATCAAAAGCTGCAACAAATCATCTTACATTTGATCCT 778066
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 CCAAAAAGTTTGAAGAAACATATGCGCTCAGAGCTGATGATTAAGAGAGTGTGATGATC 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778067 CTCAGCTGTTTGTATACACAGTATGCTATTAAGTTTCAAGTATAAAAATTCAGAAC 778126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 273 TATTCGCGAAGAAAAAAACAATATGCGCAAAACTACCAAGCTGCTTCTGCACAGCAG 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778127 TATATATAAGACAGTAAAGGCTATGATGATTAATTCATTTTCAAGCGCACTTTCACAGCAG 778186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 TATGACTCTTGAACACGTAAGCTCAATTTGTCACAGTAAATTTAGTTGAGTTGCGACT 392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778187 TTTGACTCGGAGGCTTACCAACAAATTTGCAACACTATGCTTGTGAATATGCTGT 778246
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 TAAGAAAGTAGAGAGAAGCTGATTTGACAGATGAGGCTTAAAGAAAGCTTTGATGAGTA 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778247 AAAAGAGACAGCTAAGAAAGATTTAAGACCAACTTTAAAGAGCATTAATAAAACTA 778306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 453 CACTCCAGATGTAAAGGCTCAATCATCGCTTAAATAATGAAGATTAAGGCCAAAGAGT 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778307 TACTCTGTAACCTTGTATCAAGTAACTCAATTTGATGCAAGAAAGATTAAGCTAAATCTGT 778366
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 513 TCTCGAAAAAGCCAGGCGAGAGTGTGATTTGCTCAATTAAGCCAAAGATTAATTCAGC 572
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778367 CCTTAAAGTGTAAAGGCTGATGAGCTGATTTTGCAAAAGATTGCAAAAGAAAAACAAC 778426
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 573 TGATGAAAAAACAAAAAATGCTGAGAAATTAACCTTGTATGCTTCAACAGAGAGT 632
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778427 ACCTACTGTATAGAAAGT-----GAGTATAAATTTGATTTGCGAGGGCAACACCT 778477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 633 ACCTGAAGCAAGTCAAAAAAGCCGCTTTCCCTTTAGATGTGATGCTGTTTGTGATGTAT 692
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778478 CCTTAAAGTGTAAAGGCTGATGAGCTTAAAGCTAATAAAATGCTTTCAGATGTGCT 778537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 693 TACAGCAACTGCGACACAGCCTTACAGTACCAATTTCACTTGTATGTAACCTCAATAGAA 752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778538 TTCAACGGTGTATTCACACACTTAATAAAAGTTACTACATTAATTAAGTAACTGATATA 778597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 753 AACAGAAAAATCATTAATATTTGATGACTACAAAGAAAAATTAATAAATCTTATCTTGCAC 812
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778598 GACACAGAAAAAATCTGATTTGAAATCTTACAAAATGCTTAAAGAAAGATTATCTTAA 778657
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 813 TCAAAAACAAATGATTCACACTTTGTTCAAAAGCATTTATCGAAAAAGAAATTCAGACAGC 872
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778658 GGATAAACAAAGCGATAGAGCCTTCCAAAATTAAGATTTTCAAAAGCCCTTAGAAAAAGC 778717
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 873 CAAATCAAGTTAAGAGACCAAGCCTTCCAAAATTAATCTTACCACATAT 921
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778718 TAAATGTAAGATTAAAGATTAAGCATTTGACAGCATCTTATACATATAT 778766
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
ABN70712
ID ABN70712 standard; DNA; 894 BP.
XX
XX ABN70712;
XX
XX 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide seq ID NO 9337.
XX
XX Streptococcus; GMS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX

```

```

PF 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;
XX Tettein H;
XX
XX WPI; 2002-352536/38.
XX
XX P-PsDB; ABP30081.
XX
XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
XX Claim 7; Page 4058; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 894 BP; 356 A; 146 C; 155 G; 237 T; 0 other;
XX
XX
XX Query Match
XX Best Local Similarity 26.4%; Score 255.4; DB 24; Length 894;
XX Matches 501; Conservative 0; Mismatches 356; Indels 12; Gaps 2;
XX
XX
XX 56 TCACACTATTATCAGTACGAACTTTAGCAGCTGTTTC---GAAAGGTCAGAAAGTGCAG 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TGCACTTATATGTCAGTTGCCACATTTAGCTGCTCAGGTAAACATCAATGGAACAA 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 113 ACCTTATCAGCATGAAGGGGATGTCATTAACAGACATCAATTTTATGCAAGTGAATA 172
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ATGTGTTTCAATGAAGGGCGACACTATTAAGTCTCTGATTTTATGATCAAGTAAAAA 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 173 GCAACCCCTTCAGCCCAACAAGTCTTGTAAATATGACATCCAAAAGTTTGTGANAAC 232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 CATCAAAAGCTGCACAAACAATCAATGCTTACATTTGCTCTCAGCTGTTTGTGATACAC 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 233 AATATGCTCAGAGCTTGTATGATTAAGAGAGTTGATGATTAATGCGGAAGAAAAAAC 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 AGTATGATGATTAAGAGTTGATGATTAAGAGAGTTGATGATTAATGCGGAAGTTACA 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 AATATGCGCAAAACTTACCAAGCTGCTTGTACACAGCAGTATGATCTTGAACACGTA 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 GCTATGTAATTCATTTTCAACCGCACTTTCACAAACAGGTTGATCCGGAAGTTACA 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 353 AAGCTCAATTTGTCACAACTTAATTAATTTAGTGTGCGCACTTAAAGAGTACAGAGCTG 412
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 AACACAAATTTGCGACACACTATGCTTGTGATATGCTTAAAGAGAGCTTAAGAAAG 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 413 AATTAAGAGATGAAGCTTAAAGAAAGCTTTGATGATGATGATCACTCCAGATGTAAGCGCTC 472
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 AATTAAGAGAGCAAACTTAAGAAAGAGCATATTAAGAAAGTAAAGTCTGTAAGCTGTAC 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

473 AAATCATCGCTTAAATATGAGATGAGCCAAAGATCTTCGAAAAAGCCAGGAG 532
 422 AAGTAATCAAAATTTGGATGACAGAAATGAGCTAAATCTGCTTAAAGATGTAAGGCTG 481
 533 AAGGCGTATTTTCTCAATATAGCCAAAGATTAATCACTGATGAAAAACAAAGAA 592
 482 ATGGAGCTGATTTTGCAGAAATGCAAAAAGAAAAACAACACTCTGATTAAGAAATT- 540
 593 ATGGGAGAAATATCTTTGATTTCTGCTCAACAGAAATGATACGAGCAAGTCAAAAAG 652
 541 -----GAGTATTAATTTGATTTCTGCAAGGAGCAACCTCCCTTAAGATTAAGTCA 592
 653 CCGCTTTGCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 712
 593 CAGCTTTAAGCTAATTAATAATGATGATGATGATGATGATGATGATGATGATGATG 652
 713 CTTATGAGTACCAATATATGATGATGATGATGATGATGATGATGATGATGATGAT 772
 653 CTTATGAGTACCAATATATGATGATGATGATGATGATGATGATGATGATGATGAT 712
 773 TTGATGAGTACCAATATATGATGATGATGATGATGATGATGATGATGATGATGAT 832
 713 GGAATCTTACCAAAATGCTTAAAGAAAGATTAATCTTAAGATTAAGATTAAGATG 772
 833 CATTTGTTCAAAAGCATATTCGAAAAAGATTCGAAAGCAATATCAAGGTTAAAGACC 892
 773 CTTTCAAAATTAAGTGAATTTCAAAAAGCTTGAAGAAAGCTTAATGTTAAAGATA 832
 893 AAGCTTCCCAAAATATCTTTACCCCAATAT 921
 833 AAGCATTTGACAGCATCTTATCACAATAT 861

RESULT 15
 AAX20090
 ID AAX20090 standard: DNA: 1110 BP.
 AAX20090;
 20-APR-1999 (first entry)
 Enterococcus faecalis gene EF049.
 Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic; ss.
 Enterococcus faecalis.
 WO9850554-A2.
 12-NOV-1998.
 04-MAY-1998; 98WO-US08959.
 14-NOV-1997; 97US-0066009.
 06-MAY-1997; 97US-0044031.
 16-MAY-1997; 97US-0046655.
 (HUMA-) HUMAN GENOME SCI INC.
 Bailey C, Choi GH, Hromocky J A, Kunsch CA;
 WPI, 1999-070095/06.
 P-PSDB; AAY00100.
 New isolated Enterococcus faecalis polynucleotides - used to develop
 products for the detection of Enterococcus and for use in vaccines
 for prevention or attenuation of Enterococcus infection
 Claim 1; Page 126; 301pp; English.
 The present sequence represents a gene isolated from
 Enterococcus faecalis. The present invention describes genes, proteins

CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences from the present invention can also be used for screening
 CC products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 XX
 SQ Sequence 1110 BP; 471 A; 191 C; 201 G; 244 T; 3 other;
 Query Match 22.4%; Score 216.2; DB 20; Length 1110;
 Best Local Similarity 55.4%; Pred. No. 36-40;
 Matches 513; Conservative 0; Mismatches 389; Indels 24; Gaps 4;
 28 ATGAGAAAAAATTTATGCGAGTCCGATCAACATATATCACTTACAGT 87
 82 ATGAGAAAAAATTTATCTTACCTGACGGGCGCAATGGCGTTTATGTTACAGCG 141
 88 TGTTCGAAAGGGTCAAGAGGTGACAGCTTATCAGCATGAAAGGGATGTCATACAGA 147
 142 TGTTCAGGGGTTCAAA-----AGATTCGACATGAAAGTTCAACATTTCTGTT 195
 148 CATCAATTTTATGACAGTGAAGCAACCTTACGCCAACAAGCTTGTAAATATG 207
 196 GATGATTTTATACCAATTAAGAACAAAGCACTAGCAACAAAGCTTGGCCAAATG 255
 208 ACCATCCAAAAGTTTGAAGAAATGAGTCAAGCTTATGATGATGATGATGATGAT 267
 256 GTTATTTATTAAGCTTTTGAAGAAATATGAGCAAGTATGATGATGATGATGAT 315
 268 GATGATTTGCGGAGAAAAAAGCA-----ATATGGGAAAAAATCCAGCTGTC 318
 316 AAAAATCTTGAGAGCAAGCCAAAGATGAGAGCAAGCAAGGAGGAAAGTTCTGTGCA 375
 319 TTGTCAAGCAAGGATGATGCTTGAAGAACAGTAAAGCTCAATTCGTAAGTAATTA 378
 376 TTAAAAACAAGCTGTTTAACTGAAAAAAGATTCAAGAAACAGTTAAACAAAGAGCGC 435
 379 GTTGAATGAGATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 438
 436 TATGATGAGGCTTAAA-----AGCCCATTTAAATATGATGATGATGATGAT 489
 439 GCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
 490 GCTTGGGCAAGTTTCCATCCAGAGAGAGAGCAAAATATATCCAAAGTTCTTCAGAAAT 549
 499 AAGGCCAAAGATTCTCGAAAAAGCCAGAGAGAGTGTGATTTTCTCAATTAAGCC 558
 550 GATGCCA---AAGCTGTCAAGAAAGAAATCACTGAGCGGGGATTTCAAAAATTTGCT 606
 559 AAGATTAATTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
 607 AAGAAAAATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
 619 GCTTCAAGAAAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 678
 667 CAAGCAACAAGTGTCTCCGAGTAAAGAGTGCCTTCAATTAAGATGAGCGCA 726
 679 GTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
 727 GTGTCAAGAACCAATTTGCTCAACAAATTAAGCAAACTTCAACAACTTCAATGATG 786
 739 AAGCTACTAAGAAAAACAAAAAATCATATATATGATGATGATGATGATGATGATGAT 798
 787 AAGATGACAAAAAACAAGCAAAAGGCAATGATGATGATGATGATGATGATGATGATG 846
 799 ACTGTTATCTTCACTCAAAACAAATTAATCAATTTGTTCAACATTAATGCAAAA 858
 847 AAAATGCTGAAGAAACAAATTAATGATGATGATGATGATGATGATGATGATGATG 906
 859 GAATTCAGAGCAAGCAATATCAAGTTAAGAGCAAGCTTCCAAATATCTTTACCCAA 918

Db 907 GAATTAAGCGGCCCAATGTGAATAATTAAGATGATGCCCTTCAAGACGCTTTAGCAGGC 966
 Qy 919 TATATCGGTGTGTGAGATTCAAGCTC 944
 |||||
 Db 967 TACATGCANAACCTGAATCTTCAAGCGC 992

Search completed: September 4, 2003, 19:26:26
 Job time : 325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:20:57 : Search time 87 Seconds
(without alignments)
4900.875 Million cell updates/sec

Title: US-10-049-473A-1

Perfect score: 966
Sequence: 1 agtacactatctcaaga.....gcagtagtacatcaacgaa 966

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	11309	4	US-08-961-527-108 Sequence 108, App
2	552.4	57.2	640	3	US-08-961-083-33 Sequence 33, Appl
3	552.4	57.2	640	3	US-09-536-784-33 Sequence 33, Appl
4	265.6	27.5	791	4	US-08-858-207A-219 Sequence 219, App
5	216.2	22.4	1110	4	US-09-071-035-181 Sequence 181, App
6	209.8	21.7	1035	4	US-09-107-532A-2701 Sequence 2701, Ap
7	206.4	21.4	912	4	US-09-107-532A-2459 Sequence 2459, Ap
8	203.2	21.0	970	4	US-09-071-035-183 Sequence 183, App
9	151.2	15.7	909	4	US-09-107-532A-1082 Sequence 1082, Ap
c 10	59.8	6.2	7218	1	US-08-232-463-14 Sequence 14, Appl
11	52	5.4	3095	6	5231168-1 Patent No. 5231168
12	48.8	5.1	4673	1	US-07-638-431-1 Sequence 1, Appl
13	48.8	5.1	4673	5	PCT-US92-00018-1 Sequence 578, App
14	46	4.8	633	4	US-09-134-001C-578 Sequence 1407, Ap
15	45.2	4.7	1239	4	US-09-134-001C-1407 Sequence 3, Appl
16	44.4	4.6	1891	3	US-08-973-462-3 Sequence 600, App
17	43.4	4.5	1155	4	US-09-134-001C-600 Sequence 67, Appl
18	43.4	4.5	7163	4	US-08-961-527-67 Sequence 93, Appl
19	43	4.5	835	3	US-08-961-083-93 Sequence 93, Appl
20	43	4.5	835	4	US-09-536-784-93 Sequence 93, Appl
21	43	4.5	1714	3	US-08-961-083-3 Sequence 3, Appl
22	43	4.5	1714	4	US-09-536-784-3 Sequence 3, Appl
23	43	4.5	6201	2	US-08-790-912-1 Sequence 1, Appl
c 24	43	4.5	15213	4	US-08-961-527-26 Sequence 26, Appl
25	42	4.3	1269	4	US-09-601-198-178 Sequence 178, App
26	41.6	4.3	5361	3	US-08-973-462-2 Sequence 1, Appl
27	41.6	4.3	6152	3	US-08-973-462-1 Sequence 1, Appl

28	41.4	4.3	993	4	US-09-134-001C-974 Sequence 974, App
c 29	41	4.2	10993	2	US-08-961-527-15 Sequence 15, Appl
30	40.8	4.2	3666	4	US-08-682-517-13 Sequence 13, Appl
31	40.8	4.2	3666	2	US-08-682-517-14 Sequence 14, Appl
32	40.8	4.2	4197	2	US-08-682-517-7 Sequence 7, Appl
33	40.8	4.2	4197	2	US-08-682-517-8 Sequence 8, Appl
34	40.8	4.2	4766	5	PCT-US93-07261-10 Sequence 10, Appl
c 35	40.6	4.2	729	4	US-09-601-198-16 Sequence 16, Appl
36	40.4	4.2	921	2	US-08-795-475-2 Sequence 2, Appl
37	40.4	4.2	1305	4	US-09-328-234A-6 Sequence 6, Appl
38	40.4	4.2	1308	2	US-08-795-475-4 Sequence 4, Appl
39	40.4	4.2	3222	4	US-08-714-741-39 Sequence 39, Appl
40	40	4.1	2277	1	US-08-676-967-2 Sequence 2, Appl
41	40	4.1	2277	1	US-08-676-974-2 Sequence 2, Appl
42	40	4.1	2277	2	US-09-098-487-2 Sequence 2, Appl
c 43	40	4.1	3534	4	US-09-134-001C-2269 Sequence 2269, Ap
44	39.8	4.1	1374	4	US-09-601-198-131 Sequence 131, App
45	39.6	4.1	480	4	US-09-134-001C-914 Sequence 914, App

ALIGNMENTS

```
RESULT 1
US-08-961-527-108
; Sequence 108, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage
; OPERATING SYSTEM: HP Vectra 486/33
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 108:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-108

Query Match 100.0%; Score 966; DB 4; Length 11309;
Best Local Similarity 100.0%; Pred. No. 7.4e-232;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

c 1 AGTACTCTTCTTCCTCAAGAGTAGACATGAAATAATGCGCATGCA 60
db 7632 AGTACTCTTCTTCCTCAAGAGTAGACATGAAATAATGCGCATGCA 7691
```


QY 61 CTATTATCAGTACCACTTGTAGCAGCTTGTGCAAGAGGTGACAGAGCTTATC 120
 |||||
 Db 7692 CTATTATCAGTACCACTTGTAGCAGCTTGTGCAAGAGGTGACAGAGCTTATC 7751
 |||||
 QY 121 AGCTAAGAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGCAACCT 180
 |||||
 Db 7752 AGCTAAGAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGCAACCT 7811
 |||||
 QY 181 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAGTTTAAACAAATTTGGC 240
 |||||
 Db 7812 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAGTTTAAACAAATTTGGC 7871
 |||||
 QY 241 TCAGAGCTTGATGATTAAGAGGTGATGATTAATGCGCAAGAAAAAACAATATGCG 300
 |||||
 Db 7872 TCAGAGCTTGATGATTAAGAGGTGATGATTAATGCGCAAGAAAAAACAATATGCG 7931
 |||||
 QY 301 GAAACTACCAAGTCTTGTGACAGAGGATGATGATGATGATGATGATGATGATGAT 360
 |||||
 Db 7932 GAAACTACCAAGTCTTGTGACAGAGGATGATGATGATGATGATGATGATGATGAT 7991
 |||||
 QY 361 ATTCGTACAGTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 |||||
 Db 7992 ATTCGTACAGTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8051
 |||||
 QY 421 GATGAAGCTTATGAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 |||||
 Db 8052 GATGAAGCTTATGAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8111
 |||||
 QY 481 GCTCTTAATATGAAGATGAAGGCAAGAGTTCGAAAGAGCCAGAGGATGATGATGATGATGAT 540
 |||||
 Db 8112 GCTCTTAATATGAAGATGAAGGCAAGAGTTCGAAAGAGCCAGAGGATGATGATGATGATGAT 8171
 |||||
 QY 541 GATTTTGCATTAATGACCAAGATTAATCACTGATGATGATGATGATGATGATGATGATGAT 600
 |||||
 Db 8172 GATTTTGCATTAATGACCAAGATTAATCACTGATGATGATGATGATGATGATGATGATGAT 8231
 |||||
 QY 601 GAAATTTACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 |||||
 Db 8232 GAAATTTACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8291
 |||||
 QY 661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 |||||
 Db 8292 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8351
 |||||
 QY 721 AGCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 |||||
 Db 8352 AGCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8411
 |||||
 QY 781 TACAAAGAAAAATTAATAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 |||||
 Db 8412 TACAAAGAAAAATTAATAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8471
 |||||
 QY 841 CAAGACATTTATGCAAGAAATGCAAGCAATATCAAGGTTAAGGACCAAGCCCTTC 900
 |||||
 Db 8472 CAAGACATTTATGCAAGAAATGCAAGCAATATCAAGGTTAAGGACCAAGCCCTTC 9531
 |||||
 QY 901 CAAATATCTTATGCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 |||||
 Db 8532 CAAATATCTTATGCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8591
 |||||
 QY 961 AACGAA 966
 |||||
 Db 8592 AACGAA 8597

RESULT 2 US-08-961-083-33

Sequence 33, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 33:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 640 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-083-33
 Query Match 57.2%; Score 552.4; DB 3; Length 640;
 Best Local Similarity 94.7%; Pred. No. 5,4e-129;
 Matches 593; Conservative 0; Mismatches 26; Indels 7; Gaps 2;
 QY 90 TTCGAAAGGGTCAGAGAGTGCAGACCTTATGACATGAAGAGGATGATGATGATGATGATGATGATGAT 149
 |||||
 Db 1 TTCGAAAGGGTCAGAGAGTGCAGACCTTATGACATGAAGAGGATGATGATGATGATGATGATGATGAT 60
 |||||
 QY 150 TCAATTTATGACAGATGAAAGACACCTTCAGCCCAACAGCTTTGTTAAATATGAC 209
 |||||
 Db 61 TCAATTTATGACAGATGAAAGACACCTTCAGCCCAACAGCTTTGTTAAATATGAC 120
 |||||
 QY 210 CATCCAAAAGTTTGTGAAAAACAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 269
 |||||
 Db 121 CATCCAAAAGTTTGTGAAAAACAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 |||||
 QY 270 TACTATTCCGAGAAAAAACAATATGCGGAAACATGCAAGAGTGTCTGTGACAAAGC 329
 |||||
 Db 181 TACTATTCCGAGAAAAAACAATATGCGGAAACATGCAAGAGTGTCTGTGACAAAGC 240
 |||||
 QY 330 AGGATGATCTCTTGAAGACGTAAGCTCAATTCGTAAGATTAATAGTATGATGATGATGATGATGATGAT 389
 |||||
 Db 241 AGGATGATCTCTTGAAGACGTAAGCTCAATTCGTAAGATTAATAGTATGATGATGATGATGATGATGAT 300
 |||||
 QY 390 AGTTAAGAGGTATGACAGAGCTGAATGACAGATGAAGCCATTAAGAAAGCCCTTGATGA 449
 |||||
 Db 301 AGTTAAGAGGTATGACAGAGCTGAATGACAGATGAAGCCATTAAGAAAGCCCTTGATGA 360
 |||||
 QY 450 GTACACTGCAGATGTAACGGCTCAATATCATCCGTTAATTAATGAAGATTAAGGCAAGA 509
 |||||
 Db 361 GTACACTGCAGATGTAACGGCTCAATATCATCCGTTAATTAATGAAGATTAAGGCAAGA 420
 |||||
 QY 510 AGTTCTGCAAAAACCAAGGCGAGAAGTGTGATTTGCTCAATTAAGCAAGATTAATTC 569
 |||||
 Db 421 AGTTCTGCAAAAACCAAGGCGAGAAGTGTGATTTGCTCAATTAAGCAAGATTAATTC 480
 |||||
 QY 570 AACTGATGAAAAAACAAGAAATGGTGAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 629
 |||||
 Db 481 AACTGATGAAAAAACAAGAAATGGTGAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540


```

Oy      630 AGTACGCTG--AGCAGAGTCAAAAAGCCGCTTTCGCTTAGATGTCGATGGTTCAT 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTTCGCTTTAGATGTGGATGCTTCT 600
Oy      688 G----TGATTACAGCAACTGGCACACA 710
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      601 GGATGTGATTACAGCAACTGGGCACACA 628

RESULT 3
US-09-536-784-33
: Sequence 33, Application US/09536784
: Patent No. 6573082
: GENERAL INFORMATION:
: APPLICANT: ChOI et. al.
: TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
: NUMBER OF SEQUENCES: 452
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/536,784
: FILING DATE: 30-Oct-1997
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/961,083
: FILING DATE: OCT-30-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Michelle S. Marks
: REGISTRATION NUMBER: 41,971
: REFERENCE/DOCKET NUMBER: PB340P3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 640 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-536-784-33

Query Match      57.2% Score 552.4: DB 4: Length 640:
Best Local Similarity 94.7%: Pred No. 5.4e-129:
Matches 593: Conservative 0: Mismatches 26: Indels 7: Gaps 2

Oy      90  TTCGAAAGGCTCAGAGAGGTGCAGACCTTATCAGATGAAGAAGGGAGTGTATACAGACA 149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1  TTCGAAAGGCTCAGAGAGGTGCAGACCTTATCAGATGAAGAAGGGAGTGTATACAGACA 60
Oy      150 TCAATTTTATGAGCAAGTGAAGCAACCCCTTCAGCCCAACAGATCTTGTTAAATATGAC 209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61  TCAATTTTATGAGCAAGTGAAGCAACCCCTTCAGCCCAACAGATCTTGTTAAATATGAC 120
Oy      210 CATCAAAAGATTTTGTGAAGAAACATATATGGCTCAGAGCTTGATGATTAAGAGGTGATGA 269
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 CATCAAAAGATTTTGTGAAGAAACATATATGGCTCAGAGCTTGATGATTAAGAGGTGATGA 180
Oy      270 TACTATTCCGAGAGAAAAAACATATATGGCGAAAAAGTACCAAGCTGCTTGTCCAAAGC 329
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      181 TACTATTCCGAGAGAAAAAACATATATGGCGAAAAAGTACCAAGCTGCTTGTCCAAAGC 240

```

OY	330	AGGTATGACCTCTGAAACACGTAAAGCGCAAAATTCGTCAAGTAATTAAGTGGTGGC	389
Db	241	AGGTATGACCTCTGAAACACGTAAAGCGCTCAAAATTCGTCAAGTAATTAAGTGGTGGC	300
OY	390	AGTTAAGAAGGTAGCAGAAAGCTGAAATTTGACAGATGAACCCCTATTAAGAAAGCCCTTGTATGA	449
Db	301	AGTTAAGAAGGTAGCAGAAAGCTGAAATTTGACAGATGAACCCCTATTAAGAAAGCCCTTGTATGA	360
OY	450	GTCACACTCCAGATGTACGGCTCAAAATATCCGCTTATATATGAATTAAGCCCAAGA	509
Db	361	GTCACACTCCAGATGTACGGCTCAAAATATCCGCTTATATATGAATTAAGCCCAAGA	420
OY	510	AGTTTCGCAAAAAGCCAAAGCAGAAAGGTGCTGATTTTGCATTTAGCCAAAGATAATTC	569
Db	421	AGTTTCGCAAAAAGCCAAAGCAGAAAGGTGCTGATTTTGCATTTAGCCAAAGATAATTC	480
OY	570	AACGTATGAAAAACCAAAAAGAAATGTGGAAATTAACCTTTGATTTCTGCTTCAACGA	629
Db	481	AACGTATGAAAAACCAAAAAGAAATGTGGAAATTAACCTTTGATTTCTGCTTCAACGA	540
OY	630	AGTACCTG--AGCAGCTCAAAAAGCCGCTTCCGCTTACATGTGGATGGTGTCTTGAT	687
Db	541	AGTACCTGAGCAAGTCCAAAAAACCCCTTTCCGCTTATGATGTGGATGGTGTCTT	600
OY	688	G-----TGATTAACGCAACTGGCACA	710
Db	601	GGATGTGGATTTACAGCAACTGGGGACA	628

RESULT 4
 US-08-858-207A-219
 Sequence 219, Application US/08858207A
 Patent No.: 6348328
 GENERAL INFORMATION:
 APPLICANT: Black, Michael
 APPLICANT: Hodgson, John
 APPLICANT: Knowles, David
 APPLICANT: Nicholas, Richard
 APPLICANT: Stodola, Robert
 TITLE OF INVENTION: No 6348328e1 Compounds
 NUMBER OF SEQUENCES: 552
 CORRESPONDENCE ADDRESS:
 ADDRESS: Smithkline Beecham Corporation
 STREET: 709 Swedeland Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406-0939
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/858,207A
 FILING DATE: 09-MAY-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/017670
 FILING DATE: 14-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Gimm1, Edward R
 REGISTRATION NUMBER: 38,891
 REFERENCE/DOCKET NUMBER: P50475
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-4478
 TELEFAX: 610-270-5090
 TELEX:
 INFORMATION FOR SEQ. ID NO.: 219:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 791 base pairs
 TYPE: nucleic acid
 STRADEDNESS: single

TOPOLOGY: linear
US-08-858-207A-219

Query Match 27.5%; Score 265.6; DB 4; Length 791;
Best Local Similarity 97.9%; Pred. No. 2.6e-57;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTATCTCAAGAGTAGACATGACAGAAAAAATATATGGCAGGTGCCATCACA 60
DB 507 AGTAACTATCTCAAGAGTAGACATGACAGAAAAAATATATGGCAGGTGCCATCACA 566
QY 61 CATTATAGTAGCACTTAGAGCTTGTGAAAGGGCAGAAAGCTGACAGCTTATC 120
DB 567 CATTATAGTAGCACTTAGAGCTTGTGAAAGGGCAGAAAGCTGACAGCTTATC 626
QY 121 AGCATGAAAGGGGATGTCATTCAGAACATCAATTTATGACAAAGTAAAGACACCT 180
DB 627 AGCATGAAAGGGGATGTCATTCAGAACATCAATTTATGACAAAGTAAAGACACCT 686
QY 181 TCAGCCCAACAGTCTGTTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATG 239
DB 687 TCAGCCCAACAGTCTGTTTAAATATGACCATCCAAAAGTTTGGAAAAACAATATG 746
QY 240 CTCAGAGCTTGATGATAAGAGGTGATGATCTATTGCGCGAAGA 284
DB 747 CTCAGAGCTTGATGATAAGAGGTGATGATCTATTGCGCGAAGA 791

RESULT 5

US-09-071-035-181
Sequence 181, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gail H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36, 373
REFERENCE/DOCKET NUMBER: P8369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 181:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-181

Query Match 22.4%; Score 216.2; DB 4; Length 1110;
Best Local Similarity 55.4%; Pred. No. 6.3e-45;
Matches 513; Conservative 0; Mismatches 389; Indels 24; Gaps 4;

QY 28 ATGAGAAAAAATTTATGGCAGGTGCCATCACACTATATAGTAGCACTTAGACGCT 87
DB 82 ATGAGAAAAAATTTATGGCAGGTGCCATCACACTATATAGTAGCACTTAGACGCT 141
QY 88 TGTTCGAAAGGGTCAGAGGTCAGACCTTATCAGCATGAAAGGGATGTCATTACGAA 147
DB 142 TGTTCGAAAGGGTCAGAGGTCAGACCTTATCAGCATGAAAGGGATGTCATTACGAA 195
QY 148 CATCATTTTATAGAGCAATGAAAGCAACCTTACAGCCACAAAGTCTTGTAAATATG 207
DB 196 GATGATTTTATAGCAATTAAGAAAGCAAGCACTTACAGCAAGCCGTTAGCCAAAG 255
QY 208 ACCATCCAAAGTTTGTGAAAGCAATATGCTCAGACCTTATGATATGATGATGATGAT 267
DB 256 GTTATTTAAAGTCTTTTAAAGAAAAATTTGCGACAAAGTACGACAAAGATTCAA 315
QY 268 GATPACTATTCGCGAAGAAAAAACA-----ATATGGGAAACTACCAAGCTGTC 318
DB 316 AAAAATTTTGAAGAGCCAAAGAACAGTAGAAGCAAGGCGGAAAGTCTCTGATGCA 375
QY 319 TTGTCAAGAGAGTATGACCTCTTGAACAGCTAAAGCTCAATTCGTAAGTAAATTA 378
DB 376 TTTAAACAAAGCTGTTTAACTCAAAAACATTCAGAAACAGTTAAACAAAGACGAC 435
QY 379 GTTGAAGTGGCAGTTAAGAGGTAGCAGAGCTGATGATGACAGATGAAGCCTTTAAGAA 438
DB 436 TATGATGACAGTCTTAA-----AGCCCACTTAAATTTACAGTAGAAGACTTAAACA 489
QY 439 GCCTTTGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
DB 490 GCTTGGGCAAGTTTCCATTCACAGAGTAGAGCAGCAAAATTTATCAGAGTGTCTCAAGAT 549
QY 499 AAGCCCAAGAGTCTCGAAAAAGCCAGAGGAGGAGTGTGATTTGCTCAATTTAGCC 558
DB 550 GATGCCA--AAGCTGCAAGAAAGAAATCACTAGCGGGGAGTTTCAACAAATTTGCT 606
QY 559 AAGATTAATTCACGATGAGAAAAAACAAGAAATGTTGGAGAAATTTACCTTATCT 618
DB 607 AAGAAAAATTCACGATGAGAAAAAACAAGAAATGTTGGAGAAATTTACCTTATCT 666
QY 619 GCTTCAAGAGAGTCTGAGCAAGTCAAAAAAGCGCTTTCGCTTGTGATGATGATGAT 678
DB 667 CAAGCAACAGCTGCTTCCGCAAGTAAAGAGCTGCTTCAATTTAAAGATGCGCAA 726
QY 679 GTTGTGATGATTTACAGCACTGCGACAGAGCTTACAGTACCAATTTATTTA 738
DB 727 GTTGTGATGATTTACAGCACTGCGACAGAGCTTACAGTACCAATTTATTTA 786
QY 739 AACTCAGTAAAGAAAGAAAGAAATCATCTATATGATGACTACAAAGAAATTTAAA 798
DB 787 AAGATGACGAAAGAAAGAAAGAAAGCAATGACAGAAACCTTATGAAAGAAAGATCAAG 846
QY 799 ACTGTTATCTGACTCAAAAAACAATATGATCAATTTGTTCAAGCATTTAGGAA 858
DB 847 AAAATTTGCTGAAGAAACAAATTTAGCGATCAAAATTTGTTCAAGCATTTAGTAC 906
QY 859 GAATTCGAAGCAAGCAATATCAAGGTTAAGACCAAGCTTCCAAATATCTTTACCAA 918
DB 907 GAATTTAAAGGGGCGCAATGTAAGTAAGTGTGCTTCAAGAAAGCTTTAGCAAGC 966
QY 919 TATATCGGTGGTGAAGATTTCAAGCTC 944
DB 967 TACATGCAAACTGAATCTTCAAGGCG 992

RESULT 6

US-09-107-532A-2701
Sequence 2701, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

QY	28	ATGAGCAAAAATTTATTTGGACGGTCCATCAGCTATATTATTCAGTAGCACTTTAGACGCT	87
Db	1	ATGAAAAAAAAGATTTTATAGCACTTGGCCATCTGTAAGCACTGGCTTAA---TTGTCAGGC	57
QY	88	TGTTGGAAAGGGTCAGAAAGGTGCAGACCTTATTCAGATGGAAGGGGTGTCATTACAGAA	147
Db	58	TGTAGGAATGCGCGTGAAAAAAGACCGCTT---AGTTATTAAGCGGTACATACACGGA	114
QY	148	CATTCAATTTTATGACCAAGTGAAGCAACCCCTTCAGCCCAACAGTCTTGTAAATATG	207
Db	115	CAGAAGTAATGATGATAGCTTAAAAAAAATGCAAGAGCAATTCGCCGTGCAACAGTTG	174
QY	208	ACCATCCAAAAGATTTTGAAGAAACAATATGTGGCTGCAGAGCTGATGATGATTAAGGCTGTAT	267
Db	175	ATCGTTTTCAGAGTATTTTGAAGATTAATATGAGATGATGCTATCAACCAAGAAATCGAT	234
QY	268	GATCTATTTCCGGAAGAAAAAAACAATATGGCGAAAATCCACAGCTGCTGTGCACAA	327
Db	235	TCTCATATATGATCAACCAACCAAAACAATTTGGGTGATTCATTGATTCACAAATTTGAAATCT	294
QY	328	GCAGGTATGACCTTTGAAGACGTAAGGTCCTCAATTCGTACAAGTAAATAGTTGATG	387
Db	295	GCAGGTTACACAGAACAAACATTTCAAGAGACAGTATCAACCAAGGTCCTTACAGAA	354
QY	388	GCAGTTTAAAGAGTAGAGCAGAGCTGAATATACAGATGAAGCCATATAGAAGCCCTTGTAT	447
Db	355	GGCTTAAGAA-----ACATATCAATTAACCGATGGAAGATTTTAAAAACGGCTTGGGA	408
QY	448	GAGTCACTCCAGATGTAAGCGCTCAATATCCGTCCTTAATTAATGAAAGTAAAGCCAA	507
Db	409	TCTTTCCATCCAGAAAGTGAAGGCTCAAAATTTCAAGTAGCAAGTGAAGATGATCAAAA	468
QY	508	GAAGTTTCGAAAAAGCCAGGCGAAGAGGTGCTGATTTTGTCTCAATATACCAAAATATAT	567
Db	469	GA-----TGGAAGAAAGCTGCTGATTAAGGGAGACGATTTCTCTAAATTAACCAAAAGCAAG	525
QY	568	TCAACTGATGAAAAAACAAAAAGAAATATGATGGAATTAACCTGATCTCTCAACA	627
Db	526	TCACTGATTACTAGACAAAAAAGAAAGATGGCGGAAAAAGTGAATTTGATTCCTCACTACA	585
QY	628	GAACTACTGAGCAATCAAAAAAGCCGCTTTCGCTTAAATGATGAGAGGTGTTTCTGAT	687
Db	586	ACGGTTCGGGCAAGATCTCAAGAAAGCGCTTTAAATTGAAGATGAGTCAAGTTCTGAT	645
QY	688	GTGATTAACGCAACTGGCACAAAGCCTAAGTAGGCATATTTACATGTSTAAACTGACT	747
Db	646	GTGATCACTTACCAATGCTTCACTTAATACAGGAAATATTAATGTGGTTAAAAATGGTG	705
QY	748	AAGAAAAACGAAAAATCATCTAATATTTGATGACTACACAAAGAAATTAATAAAGCTTATC	807
Db	706	AAAAAACCAAAACAAAGCAACGACATGATGAATATCAAAAAAGATTTGAAGAAATTTCA	765

DD 199 ACA

Query Match	21.0%	Score 203.2	DB 4	Length 970
Best local similarity	55.3%	Pred. No. 1.1e-41		
Matches	466	Conservative 0	Mismatches 359	Indels 18; Gaps 3
QY	111	AGACCTTATGCATGAAAGGGATGTCATTACAGAACATCAATTTTATGACCAATGAA	170	
DB	19	AGATATCCACCAATGAAAGGTTACACAAATAGTCTGATGATTTTATACCAATTTAA	78	
QY	171	AAGCAACCCCTTACGCCCAACAGTCCTGTTAAATATGACCATCCACAAAAGTTTGA	230	
DB	79	AGACAAAGCGCTAGCCACACAGCGTTTACCCAAATGTTATTTATTAAGCTTTGAA	138	
QY	231	ACAATATGGCTCAGAGCTGTATGATTAAGAAGTTGATGATACATCTTCCGAGAA	290	
DB	139	AAAATATGGCAGCAAGAACTACTGACAAAGANATTCAAAAAAAGTTTGAGCAAG	198	
QY	291	ACA-----ATATGGCGAAAACTCCAAAGCTCTCTGCACAGCAGATATGACTCT	341	
DB	199	ACAAGTAAAGACACAAAGCCGAAAGTCTCTGATGACATTTAAACAAAGCTGTTTA	258	
QY	342	TGAAAACAGTAAAGCTCAAAATTCGTACACAGTAATATAGTATGGCAGTTAAGAGGT	401	

Db 776 AATCTGAGCTAGAAAAAATGCTTACCGAAGCAAAAATCTTGATACGGAATTCATGATA 835
 QY 845 GCATTATGCAAAAGATTCGACGCAATATGCAAGTAAAGACCAACCTTCCAAA 904
 Db 836 AACAATCCGTAATGATGAAAAAGACATGTAATTAAGATCTTATGTAAAAA 895
 QY 905 ATATCTTTACCCAA 918
 Db 896 ATATATTCAAATAA 909

RESULT 10

US-08-232-463-14/C
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367

GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEFFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109

TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: PTZ9pt-F15
 US-08-232-463-14

Query Match 6.2%; Score 59.8; DB 1; Length 7218;
 Best Local Similarity 2.7%; Pred. 1.5e-05;
 Matches 10; Conservative 225; Mismatches 142; Indels 0; Gaps 0;

QY 228 AAAAAATATGCTCAGAGCTTGATGATAAAGAGTGTATGATCTATGGCGAAGAAA 287
 Db 1449 AGAAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390
 QY 288 AAAACAATATGCGCAAAACACACGCTCTGTGCACAGCAGATGACTTTGAAC 347
 Db 1389 RRR 1330

QY 348 ACCTAAGCTCAAAATTCCTACAGTAATAATGTTAGTTGGCAGTTAAGAGGTACAGA 407
 Db 1329 RRR 1270
 QY 408 AGCTGAATGACAGTGAAGCCCTATAAGAAAGCCTTGATGAGTACACTCCAGATTAAC 467
 Db 1269 RRR 1210
 QY 468 GGCTCAATCATCCGCTTAATATGATAGATAGCCCAAGAGTCTCGAAAAAGCCAA 527
 Db 1209 RRR 1150
 QY 528 GGCAGAAAGTGCATTTGCTCTCAATTAAGCAAAAGATTAATCACTGATGAAAAAACA 587
 Db 1149 RRR 1090
 QY 588 AGAAATGTGAGAAA 604
 Db 1089 RRR 1073

RESULT 11

5231168-1
 ; Patent No. 5231168
 ; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JERSEN, SOREN;
 ; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
 ; TITLE OF INVENTION: MALARIA ANTIGEN
 ; NUMBER OF SEQUENCES: 19
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/409,658
 ; FILING DATE: 18-SEP-1989
 ; SEQ ID NO: 1
 ; LENGTH: 3095
 ; 5231168-1

Query Match 5.4%; Score 52; DB 6; Length 3095;
 Best Local Similarity 4.4.7%; Pred. No. 0.00096;
 Matches 202; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 201 AATATGACCAATCCAAAAGTTTGAAGAAAACATATGCTCAGAGCTTGATGATAAGA 260
 Db 1089 AATATGACCAATTTCAAGAAATTAATGAAGATGAAGAAAGTCCACATATTCGACATGA 1148
 QY 261 GGTGATGATCTATGTTGCGAAGAAAACAAATATGCGCAAAATACCAAGCTGTCT 320
 Db 1149 AGTAGAAGTAGAAGAAATGTTCCAGAAAGATGATTAATAAAGTTGAACATGA 1208
 QY 321 GTCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTCGTACAAGTAATTA 380
 Db 1209 AGTAGAAGTTGAAGAAATTTCAACAGAAAGATTAATAAAGTCAACATGAATAGT 1268
 QY 381 TGAGTTGGCAGTTAAGAGTAGCAGAAAGTGAATGACAGATGAACCTATTAAGAAAGC 440
 Db 1269 AGAGTTGAAGAAATTTCTACAGAAAGATGATTAATAAAGTTGAACATGAATAGT 1328
 QY 441 CTTGATGATGACACTCCAGATGTAACGCTCAATATCCGCTTAATTAAGATAA 500
 Db 1329 AGAAGTTGAAGAAATTTCTACAGAAAGATTAATAAAGTTCAACATGAATAGTAGA 1388
 QY 501 GGCCAAAGAGTTCTCGAAAAAGCCAGGAGAGAGTGCTGATTTTGCTCAATTAAGCCAA 560
 Db 1389 GGTGAGAGAAATTTCTACAGAAAGATTAATAAAGTTGAACATGAATAGTAGAAGT 1448
 QY 561 AGATATTTCAAGTGAGAAAACAAAGAAATGAGTAATTAATTAAGTATTAATTAAGT 620
 Db 1449 TGAAGAAATTTCTACAGAAAGATTAATAAAGTTCAACATGAATAGTAGAGTTGA 1508
 QY 621 TTCAACAGAGTACCTGAGCAAGTCAAAAAAG 652
 Db 1509 AGAATTTCTACAGAAAGATTAATAAAG 1540

RESULT 12

Db 3697 TAAACAAAAAGTGAATATACATATTAATGAAATTCCTAAATAAATATACAAATCCAC 3756
QY 768 TATATTGATGACTACAAAGAAAATTAATACTGTTATCTTGACATCAAAAACAAATGA 827
Db 3757 CAATATATATGATTAATAAAGAAAATTAACAATAATGCTATTTTAAATTTTACAAAAAT 3816
QY 828 TTCACATTTGTTCAAAAGCATATTCGAAAGAAATTCGACAGCACAATATCAAG 883
Db 3817 AAAAATAATGCTTCTTAAGTTTATGAACTTAATAAAGTGTGATTAATAAATAATGATCG 3872

RESULT 14

US-09-134-001C-578
; Sequence 578, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ. ID NOS: 5674
; SEQ ID NO 578
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-578

Query Match 4.8%; Score 46; DB 4; Length 633;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 169; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 432 TAAAGAAAGCTTGTGATGATACACTCCAGATGTAACGGCTCAATATCCGCTCTTAATAA 491
Db 225 TAAAAAAGAGCTGAAGATATAGTTGAAAATGTAACAAAGCAAAAAGAAATTTGAAAA 284
QY 492 TGAAGTAAAGCCAAAGAAAGTTCTCGAAAAAGCCAAAGCAAGAGTGTGATTTGCTCA 551
Db 285 AGAAGGAAGGCTCTGATATATTTGAAAAAGATTTAAACAGCCAGCAATATCTTGA 344
QY 552 ATTAGCCAAAGATTAATTCATGATGAAAAACAAAAAGAAATGTGAGAAATTAACCTT 611
Db 345 ACATGTAGAAAAACAAAGCAAAAAAGAAAGTTGAACTGATAGTATTAAGA 404
QY 612 TGATCTGCTCAACAGAGTACCTGACAGATCAAAAAAGCCGCTTTCGCTTTAGATGT 671
Db 405 AAAATATTAATCAATGATGCTTATGCAAAAGCTTACAAAAAGCACTTAATTAAGAAAA 464
QY 672 GGATGCTGTTTCTGATGATTAACAGCACTGGACACAGAGCTACAGTACAGCAATATTA 731
Db 465 AGAAGCTGTTTCTTATTAATGAAGATATGCAACATCGGAAGTAGAGGAAAAATC 524
QY 732 CATTTAAACCTACATAGAAAAAGAAAAATCATCTAATTAATGATGACTACAAAGAAA 791
Db 525 GAAAGATCTTTCTTAAGCATATTAAGAAATGAATTAATTAATTAATGCTTACTCAAAAGC 584
QY 792 ATTAATAACTGTTA 805
Db 585 CATGAGAAAGTTAA 598

RESULT 15

US-09-134-001C-1407
; Sequence 1407, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1407
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1407

Query Match 4.7%; Score 45.2; DB 4; Length 1239;
Best Local Similarity 45.3%; Pred. No. 0.034;
Matches 164; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 243 AGAGCTTGAATGAAGAAGTGTATGATATCTATTCGCCGAAGAAAAACAAATATGCGCA 302
Db 837 AGATTATCAAAATCAAGTCAATTAATAATGCAAGGCTTTAGCTCAAAATTAATTAAGA 896
QY 303 AAACACCAACGTGTCTTGTCAACAGCAGTATGACTCTTGAACACGTAAGCTCAAT 362
Db 897 AGGCTTCAGAGTGTATCTGCGCGGAAACAGCATCATTTAGTGCAGTTGATCTTAAAG 956
QY 363 TCGTACAAATTAATTTGATGAGTTGGCAGTTTAAAGAGTGAAGCTGAATTCACAGA 422
Db 957 TTCTATCAATATGACTGCTGTAATTAAGCTGAAAGACACTTGATTAAGTGAATTCATCG 1016
QY 423 TGAAGCCTTAAGAAGACCTTGTGATGATACACTCCAGATGTAACGGCTCAATATCCG 482
Db 1017 TATATAGATATGATATTCATTTGATTAAGAAAAACCTTTGTAACAGTGTGATGACT 1076
QY 483 TCTTAATTAATGAAGATTAAGCCAAAGAGTTCTGAAAAAGCCAAAGCAAGAGTCTGA 542
Db 1077 AGGAAACCCAGCAGCTCAACACTGCTGTTTGTGATGATCTGCTTTGTGGAAGTAGCAAA 1136
QY 543 TTTTGCTCAATTAGCCAAAGATTAATTCACACTGATGAAAAAGAAAGAAATGCTGAGA 602
Db 1137 AATTATTAATTTAGCTTAAATTAATGATATGATTAATTAATTAATTAATTAATTAAG 1196
QY 603 AA 604
Db 1197 GA 1198

Search completed: September 4, 2003, 21:03:46
Job time : 90 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 21:02:12 ; Search time 315 Seconds
(without alignments)
7053.264 Million cell updates/sec

Title: US-10-049-473a-1
Perfect score: 966
Sequence: 1 agtaacctatctcaag.....gcagtagcatcaacga 966

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1537136 seqs, 114998732 residues
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PTCT_US_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	939	97.2	942	11	US-09-769-787-179
2	552.4	57.2	640	9	US-09-765-272-33
3	215.6	22.3	12860	10	US-09-070-927A-144
4	47.6	4.9	354	13	US-10-101-487-69
5	47.2	4.9	554	13	US-10-101-487-106
6	46.2	4.8	522	13	US-10-101-487-71
7	46.2	4.8	530	13	US-10-101-487-73
8	46	4.8	6189	12	US-10-240-485-145
9	46	4.8	6912	12	US-10-240-485-246
10	45.6	4.7	15674	12	US-10-311-455-336
11	45.6	4.7	15674	12	US-10-240-485-30
12	45.4	4.7	7104	9	US-09-815-242-4580
13	45.4	4.7	7107	9	US-09-815-242-8291
14	45.4	4.7	9964	12	US-10-311-455-71
15	45	4.7	2142	10	US-09-917-800A-1567
16	45	4.7	3042	10	US-09-870-759-85

17	45	4.7	3042	12	US-09-751-708A-85	Sequence 85, Appl
18	45	4.7	6389	12	US-10-311-455-1934	Sequence 1934, Ap
19	45	4.7	7624	12	US-10-311-455-1953	Sequence 1953, Ap
20	44.8	4.6	17331	12	US-10-311-455-1026	Sequence 1026, Ap
21	44.4	4.6	1891	10	US-09-742-096-4	Sequence 4, Appl
22	44.4	4.6	5467	12	US-10-311-455-2082	Sequence 2082, Ap
23	44.2	4.6	225	10	US-09-764-868-1268	Sequence 1268, Ap
24	44.2	4.6	5336	12	US-10-311-455-2072	Sequence 2072, Ap
25	44.2	4.6	6685	12	US-10-311-455-879	Sequence 879, App
26	44.2	4.6	9686	12	US-10-311-455-1662	Sequence 1662, Ap
27	43.8	4.5	567	10	US-09-764-868-24	Sequence 24, Appl
28	43.8	4.5	567	11	US-09-955-999-31	Sequence 31, Appl
29	43.6	4.5	433	10	US-09-960-352-3277	Sequence 3277, Ap
30	43.6	4.5	6314	12	US-10-240-453-14	Sequence 14, Appl
31	43.6	4.5	8876	12	US-10-311-455-2050	Sequence 2050, Ap
32	43.4	4.5	3231	12	US-10-311-455-1262	Sequence 1262, Ap
33	43.4	4.5	5508	12	US-10-311-455-1372	Sequence 1372, Ap
34	43.4	4.5	6174	12	US-10-311-455-1482	Sequence 1482, Ap
35	43.2	4.5	7195	12	US-10-240-453-40	Sequence 40, Appl
36	43.2	4.5	7195	14	US-10-239-676-30	Sequence 30, Appl
37	43	4.5	835	9	US-09-765-272-93	Sequence 93, Appl
38	43	4.5	1714	9	US-09-765-272-3	Sequence 3, Appl
39	43	4.5	5179	12	US-10-311-455-1969	Sequence 1969, Ap
40	43	4.5	5179	12	US-10-240-453-277	Sequence 277, App
41	43	4.5	7255	12	US-10-311-455-1251	Sequence 1251, Ap
42	43	4.5	9483	12	US-10-311-455-349	Sequence 349, App
43	43	4.5	14649	12	US-10-240-453-142	Sequence 142, App
44	43	4.5	14649	14	US-10-239-676-122	Sequence 122, App
45	43	4.5	15416	12	US-10-311-455-2203	Sequence 2203, Ap

ALIGNMENTS

US-09-769-787-179	RESULT 1
Sequence 179, Application US/09769787	
Publication No. US20030091577A1	
GENERAL INFORMATION:	
APPLICANT: Microbial Technics Limited	
APPLICANT: Gilbert, Christophe	
APPLICANT: Hansbro, Philip M	
TITLE OF INVENTION: Proteins	
FILE REFERENCE: PWC/P21129NO	
CURRENT APPLICATION NUMBER: US/09/769,787	
CURRENT FILING DATE: 2001-01-26	
PRIOR APPLICATION NUMBER: GB 9816337.1	
PRIOR FILING DATE: 1998-03-27	
PRIOR APPLICATION NUMBER: US 60/125164	
PRIOR FILING DATE: 1999-03-19	
NUMBER OF SEQ ID NOS: 388	
SOFTWARE: Patentln Ver. 2.1	
SEQ ID NO 179	
LENGTH: 942	
TYPE: DNA	
ORGANISM: Streptococcus pneumoniae	
US-09-769-787-179	
Query Match	97.2% Score 939 DB 11: Length 942:
Best Local Similarity 100.0% Pred. No. 3.7e-190:	
Matches 939: Conservative 0: Mismatches 0: Indels 0: Gaps 0:	
28 ATGAAGAAAAAATTTATGGCAGGTGCATCACACTTATACATAGCAACTTACAGACT 87	
1 ATGAAGAAAAAATTTATGGCAGGTGCATCACACTTATACATAGCAACTTACAGACT 60	
88 TGTTCGAAGAGGTGCAAGGTGCAGACTTATACAGATGAAGGGATGTCATTACAGAA 147	
61 TGTTCGAAGAGGTGCAAGGTGCAGACTTATACAGATGAAGGGATGTCATTACAGAA 120	
148 CATCAATTTTATAGCAAGTGAAGCAACCCCTTCAGCCCAACAATCTGTTAAATFAG 207	
121 CATCAATTTTATAGCAAGTGAAGCAACCCCTTCAGCCCAACAATCTGTTAAATFAG 180	

```

QY 208 ACCATCCAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTGAT 267
   |||||||
Db 181 ACCATCCAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTGAT 240
QY 268 GATGATATGCGGAGAGAAAAAACAATATGCGGAGAAACTACCAAGCTGCTGTGCACA 327
   |||||||
Db 241 GATGATATGCGGAGAGAAAAAACAATATGCGGAGAAACTACCAAGCTGCTGTGCACA 300
QY 328 GAGGATATGCTCTTGAACACAGTAAAGCTCAATTCGTCACAGTAATTAAGTGTG 387
   |||||||
Db 301 GAGGATATGCTCTTGAACACAGTAAAGCTCAATTCGTCACAGTAATTAAGTGTG 360
QY 388 GAGGATTAAGAGTACAGAGAGCTGTAATGACAGATGAGAGCTTAATGAAGAGCTTTGAT 447
   |||||||
Db 361 GAGGATTAAGAGTACAGAGAGCTGTAATGACAGATGAGAGCTTAATGAAGAGCTTTGAT 420
QY 448 GAGTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATTAATGAGATTAAGGCCAAA 507
   |||||||
Db 421 GAGTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATTAATGAGATTAAGGCCAAA 480
QY 508 GAGGATTCGAAAAAGCCAGAGAGAGGCTGATTTGCTCAATTAAGCCAAAGTAAAT 567
   |||||||
Db 481 GAGGATTCGAAAAAGCCAGAGAGAGGCTGATTTGCTCAATTAAGCCAAAGTAAAT 540
QY 568 TCAACTGATGAAAAAACAAGAAAAATGCTGAGAAATTAACCTTTGCTGCTTCAACA 627
   |||||||
Db 541 TCAACTGATGAAAAAACAAGAAAAATGCTGAGAAATTAACCTTTGCTGCTTCAACA 600
QY 628 GAAGTACCTGAGACAGTCAAAAAAGCCGCTTTCGCTTTAGATGATGATGATGAT 687
   |||||||
Db 601 GAAGTACCTGAGACAGTCAAAAAAGCCGCTTTCGCTTTAGATGATGATGATGATGAT 660
QY 688 GTGATTAACAGCAATGCGACACAGAGCTTACAGTAAATTAATTAATTAATTAAT 747
   |||||||
Db 661 GTGATTAACAGCAATGCGACACAGAGCTTACAGTAAATTAATTAATTAATTAAT 720
QY 748 AAGAAAAAGAAAAATCATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 807
   |||||||
Db 721 AAGAAAAAGAAAAATCATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
QY 808 TTGACATCAAAAAACAATGATGATCAATTTGTCAAAGCTTAATGAAAAAGATGCAA 867
   |||||||
Db 781 TTGACATCAAAAAACAATGATGATCAATTTGTCAAAGCTTAATGAAAAAGATGCAA 840
QY 868 GCAGCAATATCAAGGTTAAGACCAAGCCCTTCCAAAATATCTTAACCAATATATCGGT 927
   |||||||
Db 841 GCAGCAATATCAAGGTTAAGACCAAGCCCTTCCAAAATATCTTAACCAATATATCGGT 900
QY 928 GTGAGATTCAGAGCTCAAGCACTAGTACATCAAAAGCAA 966
   |||||||
Db 901 GTGAGATTCAGAGCTCAAGCACTAGTACATCAAAAGCAA 939

```

RESULT 2
US-09-765-272-33

Sequence 33, Application US/09765272
Patent No. US200200615451

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

City: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

```

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765, 272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-765-272-33

Query Match      57.2%; Score:552.4; DB 9; Length 640;
Best Local Similarity 94.7%; Pred. No.5,4e-108;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTGAAAAGGCTGAGAAAGTGCAGACCTTATTCAGATGAAGAGGAGATGTCATTACAGACA 149
   |||||||
Db 1 TTGAAAAGGCTGAGAAAGTGCAGACCTTATTCAGATGAAGAGGAGATGTCATTACAGACA 60
QY 150 TCAATTTATGAGCAAGTGAAGCAACCCTTACGCCCAAGTCTGTTAAATATGAC 209
   |||||||
Db 61 TCAATTTATGAGCAAGTGAAGCAACCCTTACGCCCAAGTCTGTTAAATATGAC 120
QY 210 CATCCAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTTGATGA 269
   |||||||
Db 121 CATCCAAAAGTTTGAACAAATATGCTCAGAGCTTGATGATTAAGAGGTTGATGA 180
QY 270 TACTATTGCGGAGAAAAAACAATATGCGGAGAAACTACCAAGCTGTCTTGCACAGC 329
   |||||||
Db 181 TACTATTGCGGAGAAAAAACAATATGCGGAGAAACTACCAAGCTGTCTTGCACAGC 240
QY 330 AGGTATGACTCTTGAACACGTTAAAGCTCAATTCGTAAAGTAATTAAGTGTGAGTTGGC 389
   |||||||
Db 241 AGGTATGACTCTTGAACACGTTAAAGCTCAATTCGTAAAGTAATTAAGTGTGAGTTGGC 300
QY 390 AGTTAAGAAGTGAAGAGAGCTGAATGACAGATGAAGAGCTTAATGAAGAGCTTTGATGA 449
   |||||||
Db 301 AGTTAAGAAGTGAAGAGAGCTGAATGACAGATGAAGAGCTTAATGAAGAGCTTTGATGA 360
QY 450 GTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATTAATGAAGATTAAGGCCAAAAG 509
   |||||||
Db 361 GTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATTAATGAAGATTAAGGCCAAAAG 420
QY 510 AGTTCTCGAAAAAGCCAGAGAGAGAGTGTGATTTGTCATTAATGCAATTAAGATTAATTC 569
   |||||||
Db 421 AGTTCTCGAAAAAGCCAGAGAGAGAGTGTGATTTGTCATTAATGCAATTAAGATTAATTC 480
QY 570 AACTGATGAAAAAACAAGAAAAATGTTGAGAGAAATTAACCTTTGATTTGCTTCAACAGA 629
   |||||||
Db 481 AACTGATGAAAAAACAAGAAAAATGTTGAGAGAAATTAACCTTTGATTTGCTTCAACAGA 540
QY 630 AGTACCTG--AGCAAGTCAAAAAAGCCGCTTTCGCTTTAAGATGATGATGATGATGAT 687
   |||||||
Db 541 AGTACCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAAGATGATGATGATGATGATGAT 600
QY 688 G-----TGATTACGCACTGGCAGACA 710
   |||||||
Db 601 GGATGTGATTTACAGCAAACTGGGGCACA 628

```


NAME/KEY: CDS
LOCATION: (1)..(528)
US-10-101-487-69

Query Match 4.9%; Score 47.6; DB 13; Length 554;
Best Local Similarity 44.3%; Pred. No. 1.2;
Matches 194; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 215 AAAAGTTTGAACAAATATGCTCAGACCTTGATGATTAAGAGTGTGATGATCA 274
DB 2 ATAAATATCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
QY 275 TTGCCGAGAAAAAACAATATGCGCAAACTACCAAGTCTTGTCTACAAGCAGTA 334
DB 62 AGGAG 121
QY 335 TGACTCTTGAACACGTAAAGCTCAATTCGTACAGTAATTAAGTTGAGTTGCA 394
DB 122 AGCAGGAG 181
QY 395 AGAAGTAGCAG 454
DB 182 AGGAG 241
QY 455 CTCGAGATGTAACGGCTCAATTCCTTAAATTAAGAGAGAGAGAGAGAGAG 514
DB 242 AGCAGGAG 301
QY 515 TGGAAG 574
DB 302 AGGAG 361
QY 575 ATGAAAAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
DB 362 AGGAG 421
QY 635 CTGAGCAAGTCAAAAAG 652
DB 422 AGGAG 439

RESULT 5

US-10-101-487-106
Sequence 106, Application US/10101487
Publication No. US20020169125A1
GENERAL INFORMATION:
APPLICANT: LEUNG, DAVID W.
APPLICANT: BERGMAN, PHILIP A.
APPLICANT: LOFOULIST, ALAN
APPLICANT: PIETZ, GREGORY E.
APPLICANT: TOMPKINS, CHRISTOPHER K.
APPLICANT: WAGGONER JR., DAVID W.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
FILE REFERENCE: 077319/0329
CURRENT APPLICATION NUMBER: US/10/101,487
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/277,705
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 106
LENGTH: 554
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: oligonucleotide
NAME/KEY: CDS
LOCATION: (1)..(537)
US-10-101-487-106

Query Match 4.9%; Score 47.2; DB 13; Length 554;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 190; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 225 TGAAAAACAATATGCTCAGAGCTTTGATGATTAAGAGTGTATGATCTATGCGGAGAG 284
DB 12 TGAAG 71
QY 285 AAAAACAATATGCGCAAACTACCAAGTCTTGTCTACAAGCAGAGTATGACTCTTGA 344
DB 72 AGAAG 131
QY 345 AACAGCTAAAGCTCAATTCCTTACAGTAATTAAGTTGAGTTGAGCTTAAGAGCTT 404
DB 132 AGAAG 191
QY 405 AGAAGCTAATTTGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
DB 192 AGAAG 251
QY 465 AACGCTCAATTCATCCCTCTTAATTAAGATTAAGAGAGAGAGAGAGAGAGAG 524
DB 232 AGAAG 311
QY 525 CAGGCAAG 584
DB 312 AGAAG 371
QY 585 AAAAAG 644
DB 372 AGAAG 431
QY 645 CAAAAAG 652
DB 432 AGAAGAGAG 439

RESULT 6

US-10-101-487-71
Sequence 71, Application US/10101487
Publication No. US20020169125A1
GENERAL INFORMATION:
APPLICANT: LEUNG, DAVID W.
APPLICANT: BERGMAN, PHILIP A.
APPLICANT: LOFOULIST, ALAN
APPLICANT: PIETZ, GREGORY E.
APPLICANT: TOMPKINS, CHRISTOPHER K.
APPLICANT: WAGGONER JR., DAVID W.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
FILE REFERENCE: 077319/0329
CURRENT APPLICATION NUMBER: US/10/101,487
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/277,705
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 71
LENGTH: 522
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: oligonucleotide
NAME/KEY: CDS
LOCATION: (1)..(522)
US-10-101-487-71

Query Match 4.8%; Score 46.2; DB 13; Length 522;
Best Local Similarity 44.3%; Pred. No. 2.3; Mismatches 238; Indels 0; Gaps 0;
Matches 189; Conservative 0;

OY	226	GAATAACAAATATGCTCAGAGCTTGATGATTAAGAGCTTGATGATCTATTGGCGAAGAA	285
Db	1	GAAGCGAAGAGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA	60
OY	286	AAAAAACAAATATGCGAAAAATACCAACGCTGCTCTTGTCACAAGCAGTATGATCTTTGAA	345
Db	61	GAAGGGAAGAGGAAGAA	120
OY	346	ACACGTAAGCTCAAATTCGTAACAAGTAATTAGTTGAGTTGGCAGTTTAAGAAAGTAGCA	405
Db	121	GAAGGGAAGAGGAAGAA	180
OY	406	GAACTGAAATTCAGCAGATGAACCTATATAAGAAAGCTTTGATGAGTAGACATCCAGATGTA	465
Db	181	GAAGAGGAAGAGGAAGAA	240
OY	466	ACGCGTCAAAATCATCCGCTTTATATATGAGATTAAGGCCAAGAAAGCTTCGAAAAAGCC	525
Db	241	GAAGGGAAGAGGAAGAA	300
OY	526	AAGCGAGAAGAGGTCGATTTTGCTCAATTAGCCAAAGATAAATTCACATGATGAAAAACA	585
Db	301	GAAGCGAAGAGGAAGAA	360
OY	586	AAAGAAAATGGTGAAGAAATTAACCTTGTATTCGCTTCAACAGAACTACCTGAGCAATGC	645
Db	361	GAAGGGAAGAGGAAGAA	420
OY	646	AAAAAAG 652	
Db	421	GAAGAGG 427	

RESULT 7
 US-10-101-487-73/c
 Sequence 73, Application US/10101487
 Publication No. US20020169125a1
 GENERAL INFORMATION:
 APPLICANT: LEUNG, DAVID W.
 APPLICANT: BERGMAN, PHILIP A.
 APPLICANT: LOFOUST, ALAN
 APPLICANT: PIETZ, GREGORY E.
 APPLICANT: TOMPKINS, CHRISTOPHER K.
 APPLICANT: MAGGONER JR., DAVID W.
 TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
 TITLE OF INVENTION: THEREOF
 FILE REFERENCE: 077319/0329
 CURRENT APPLICATION NUMBER: US/10/101,487
 CURRENT FILING DATE: 2002-03-20
 PRIOR APPLICATION NUMBER: 60/277,705
 PRIOR FILING DATE: 2001-03-21
 NUMBER OF SEQ ID NOS: 116
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 73
 LENGTH: 530
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 US-10-101-487-73

	Query Match	Best Local Similarity	Score 46.2%	DB 13	Length 530:
	Matches	189:	Conservative	0:	Mismatches 239: Indels 0: Gaps 0:
OY	226	GAATAACAAATATGCGCTCAGACCTTGATATATATAAGAGGTGGATGATACATATTTGCCGAAGA	285		
DB	530	GAGAGCGAAGAGGAAGAAGAGAGAGAGAGAAAGAGGAAGGAAGGAAGAGAGAGAGGAAGAA	471		
OY	286	AAAAAACATATATGCGCAAACTACCAACGTTCTTTGCACAAGCAGATGATGACCTTTGAA	345		
DB	470	GAAAGCGAAGAGGAAGAAGACAGAGAAAGAAAGGAAGGAAGGAAGAGAGAGGAAGAA	411		

Oy	346	ACACGTAAAGCCAAATTCGTCACAGTAATAATTAGTTGATGGCATGGCATTAAGAAGTAGCA	405
Oy	410	GAACAGGAAGAGAGAAAGAA	351
Oy	406	GAACGTGAATTGCAGATGAACCTATAGAAGACCTTTGATGAGTACACTCCAGATGTA	465
Oy	350	GAAGAGGAAGAA	291
Oy	466	ACGGCTCAATCATCCGTCTTAAATATGAAGATTAAGGCCAAAGAAAGTTCTCGAAANAGCC	525
Oy	290	GAAGAGGAAGAA	231
Oy	526	AAGCCAGAAGGTGCTGATTTTGGCTCAATTAGCCAAAGATTAATTCATGATGAAAAAACA	585
Oy	230	GAAGAGGAAGAA	171
Oy	586	AAAGAAATGCGTGAATAATTAACCTTTGATTTCTGCTTCAACAGAGATACCTAGCANATC	645
Oy	170	GAAGAGGAAGAA	111
Oy	646	AAAAAAG 652	
Oy	110	GAAGAGG 104	

```

RESULT 8
US-10-240-485-145/C
; Sequence 145, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240.485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 145
; LENGTH: 6189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-145

```

	Query Match	4.8%: Best Local Similarity	Score 46; 50.0%: Pred. No. 7;	DB 12; Mismatches 115; Indels 0; Gaps 0;	Length 6189;
Qy	692	TTACAGCAGCTGGACACACAGCCCTACAGTACCAATATTACATTGTAAAACTCACTAGA	751		
Db	3418	TAAATTCATTAACATTATTACTGTCGTCATTTTCCAAAAATTAACCTAAATTCACACACAAAA	3359		
Qy	752	AAACGAAAAAATCATCTAATATTGATGACCTACAAAGAAAAAATTAAAAAAGCTGTATCTTGA	811		
Db	3358	AAAAAAAACCATTAATTCATTAATTAACCTAAAAATATACTAAAAAATAATATAAAACTA	3299		
Qy	812	CTCAAAAAACAATATGATTCAACATTTGTTCCAAGCATTTATGGGAAAGCAATTCGACAGCAG	871		
Db	3298	TTCTACTACAAATAATCGTAAATTAACATTTTAAATTAAGTCTCAAAAAATTAATAAAAAATAAA	3239		

```

Oy      872  CCAATATCAAGGTTAAGGACCAAGCCTTCACAAAATATCTTTACCCAAATAT  921
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3238 ACATTTACACATACACATCTATACCTTAACACACATTTCTCTCAATAT  3189

```

RESULT 9
US-10-240-453-246/c
; Sequence 246, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-246

	Query Match	4.8%	Score 46;	DB 12;	Length 6912;
	Best Local Similarity	50.5%;	Pred. No. 7.4;		
	Matches	112;	Conservative	0;	Mismatches 110; Indels 0; Gaps 0;
QY	659	AACGTGGCACACAGCCTACAGTACGCATATTACITTTAAACCTACTAAGAAACAGA	758		
	111	111	111	111	111
DB	1573	AACCTTAACCCCTAATATTAACAAAAACCTTAATATTAACCAACAAATACAAATATAA	1514		
QY	759	AAATCATCTATATATGTGATGACTACAAAGAAAAATTAACCTGATCTTGACTCAAAA	818		
	111	111	111	111	111
DB	1513	AAAAACAAAAAAACGATTAATATTAATATTAATTAATTAATTAATTAATTAATTAATTAAT	1455		
QY	819	ACAAATGATTCACATTTGTTCAAGCATTTTCGGAAGAAAGATTCAGACGCCAATAT	878		
	111	111	111	111	111
DB	1453	AAAAATATATCTTAATTAATTTTCATTAATAAATTAATTAATTAATTAATTAATTAATTAAT	1394		
QY	879	CAAGTTAAGGACCAAGCCTCCAAATATCTTTACCCCAATA	920		
	111	111	111	111	111
DB	1393	TATATATTAATTAACCAATATCATTAATTTTCACCACTTAA	1352		

```

: RESULT 10
: US-10-311-455-336/c
: Sequence 336, Application US/10311455
: Publication No. US20030143606A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of diseases Associated with the Immune System by Determining Cytosine Methylation
: TITLE OF INVENTION: Cytosine methylation
: FILE REFERENCE: 5013.1014
: CURRENT APPLICATION NUMBER: US/10/311,455

```

```

: CURRENT FILING DATE: 2002-12-16
: PRIOR APPLICATION NUMBER: PCT/EP01/07537
: PRIOR FILING DATE: 2001-07-02
: PRIOR APPLICATION NUMBER: DE 10032529.7
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: DE 10043826.1
: PRIOR FILING DATE: 2000-09-01
: NUMBER OF SEQ ID NOS: 2424
: SEQ ID NO 336
: LENGTH: 13674
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-336

```

Query Match	4.78;	Score 45.6;	DB 12;	Length 15674;
Best Local Similarity	45.78;	Pred. No. 13;		
Matches 159;	Conservative 0;	Mismatches 169;	Indels 0;	Gaps 0

[illegible]

```

; RESULT 11
; US-10-240-485-30/c
; Sequence 30, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10033529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 30
; LENGTH: 15674
; TYPE: DNA
; ORGANISM: Artificial Sequence

```


DB 6722 AAGCGACTTCGACTGAGATAGTCACAAAAGCTGACATGCGACACAGAGATACCAAGATTA 6781
DB 386 TGGACGTTAAGAGTAGAGAGAAAGCTGAATTTGACATGACATGAAGCTTTAAGAAAGCTTTG 445
DB 6782 ATCAACGATCAATTTGGTGCACACTGACAGATGTAAACATAGCTACGCAAGAGAGGGTG 6841
DB 446 ATGAGTACACTCCAGATGTAGCGCTCAATCATCCGCTCTTAATTAATGAAGATAAGGCA 505
DB 6842 CCAATGATCTCCAGCTACAGTTTCAAAAAGAGCAATAGTGTATATCAAGATATGCTTA 6901
DB 506 AAGAGTCTTCGAAAAAGCCAAAGGCAAGAGTGTGATTTTGTCTCAATTAGCCAAAGATA 565
DB 6902 AGTACTAAGTAAAGAAACCAAGCAAAATGCAAAATGAGCCCAAGCAAGAAAGTTA 6961
DB 566 ATTCAACTGATGAAAAACAAA 588
DB 6962 ATTAGCTTAACACACAAAGCTAAA 6984

RESULT 14

US-10-311-455-71/C
; Sequence: 71, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 71
; LENGTH: 9964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

Query Match 4.7%; Score 45.4; DB 12; Length 9964;

Best local Similarity 42.2%; Pred. No. 11;

Matches 331; Conservative 0; Mismatches 446; Indels 8; Gaps 1;

DB 139 ATTACGAACATCAATTTATATGAGCAAGTGAAGAACCCCTTCAGCCCAACAGAGCTTG 198
DB 9131 ATAAACAAATCTAAATATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 9072
DB 199 TTAATATGACATCCAAAAGTTTGTGAATAAATATGCTCAGAGCTTATGTATAA 258
DB 9071 AAATTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 9012
DB 259 GAGTTGATGATCACTATTCGCGAAGAAAAAACAATATGCGGCAAACTACCAAGCTGTC 318
DB 9011 AAAATTAATAAAACAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 8952
DB 319 TTGTCAACAAGAGTATGACTCTTGAACACGTAAGCTCAATTCGTACAAAGTAATAA 378
DB 8951 AAAATTAATAAAACAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 8892
DB 379 GTTGAAGTGGCACTTAAGAAAGTAGAGAGAGCTGAATTTACAGATTAAGCTATAGAAA 438
DB 8891 AAATATAATAATATACACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 8832
DB 439 GCCTTGTATGATGATACACTCCAGATGTAAAGGCTCAAAATCATCGCTTAAATATAGAAAT 498

DB 8831 AAACATTAATAAAAAAATCAATTAATCTTAATAAATTTATACAAAATTCATATAAATAA 8772
DB 499 AAGGCAAAAGAGTTTCCAAAAAGCCAAAGCAAGAGTGTGATTTTCTCAATTAAGCC 558
DB 8771 AAAAATTAATAAATAATTCACAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8720
DB 559 AAAAGATATTCACATGATGAAAAAACAAGAAATGTTGAGAAATTAATCTTTGATTCCT 618
DB 8719 AAAAATTAATCAAAAAAATTAATCAAAAAAATTAATTAATTAATTAATTAATTAATTAAT 8660
DB 619 GCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCCTTTGCTTTAGATGTGATGCT 678
DB 8659 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8600
DB 679 GTTTCGATGTGATTCACAGCACTGCAACAGAGCTTAAGTACCAATTAATTAATTAAT 738
DB 8599 AATCAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8540
DB 739 AAACCTACTAAGAAAAACAGAAAAATCATCTAATTAATGATGACTACAAAGAAAAATTA 798
DB 8539 AAACCTAATAAACACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8480
DB 799 ACTGTATCTTGACTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 858
DB 8479 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8420
DB 859 GAATTCAGAGCCCAATATCAAGTTAAGACCAAGCTTCCAAATATCTTTACCCCA 918
DB 8419 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8360
DB 919 TATAT 923
DB 8359 TAAAT 8355

RESULT 15

US-09-917-800A-1567
; Sequence 1567, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1567
; LENGTH: 2142

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:14:57 ; Search time 2071 Seconds

(without alignments)
11336.612 Million cell updates/sec

Title: US-10-049-473a-1

Perfect score: 966

Sequence: 1 agtaacctatctcaag.....gcagtagatcataaagaa 966

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estbha:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	72.4	7.5	868	28	BH161951
C 2	71.6	7.4	843	28	AZ551618
C 3	71.6	7.4	931	28	BH160272
C 4	71	7.3	877	28	AZ531291

5	71	7.3	908	28	AZ548467	AZ548467	ENTEK30TR
6	71	7.3	912	28	AZ551092	AZ551092	ENTFJ22TF
7	69.8	7.2	905	28	AZ550256	AZ550256	ENTFV58TR
8	66.6	6.9	849	28	AZ546009	AZ546009	ENTFW53TF
9	66	6.8	938	28	AZ676392	AZ676392	ENTHY47TR
10	64.8	6.7	890	28	AZ530768	AZ530768	ENTBH54TF
11	64.6	6.7	976	28	BH149983	BH149983	ENTOD93TF
12	64.2	6.6	880	28	AZ669474	AZ669474	ENTYR88TR
13	63.2	6.5	665	13	BG432980	BG432980	603221001
14	62.8	6.5	865	28	AZ529191	AZ529191	ENTBV68TR
15	62.8	6.5	881	28	BW266138	BW266138	BW266138
16	61.6	6.4	677	12	BH161747	BH161747	EST564270
17	60.8	6.3	1101	29	CNS0039G	CNS0039G	AL063921
18	60.4	6.3	532	12	BM273998	BM273998	EST564270
19	60.4	6.3	719	14	CA855416	CA855416	PFEST0A6
20	60.2	6.2	816	28	AZ535744	AZ535744	ENTQ25TR
21	60	6.2	890	28	BH146886	BH146886	ENTPK48TF
22	59.8	6.2	879	28	AZ550718	AZ550718	ENTEM36TF
23	59.8	6.2	942	28	BH148582	BH148582	ENTPR62TF
24	59.2	6.1	656	13	BH144040	BH144040	BM144040
25	59.2	6.1	721	13	BM141179	BM141179	BM141179
26	59.2	6.1	891	28	AZ683582	AZ683582	ENTKK47TR
27	59.2	6.1	942	14	CD050127	CD050127	AGENC00RT
28	59	6.1	906	28	BH153606	BH153606	ENTTS83TF
29	58.8	6.1	900	28	AZ549980	AZ549980	ENTTD94TF
30	57.4	5.9	748	29	BX137558	BX137558	Danio rer
31	57	5.9	671	13	BM273704	BM273704	BM273704
32	57	5.9	727	29	BX159538	BX159538	Danio rer
33	57	5.9	1061	14	CD386766	CD386766	AGENC00RT
34	56.8	5.9	1184	28	B13117	B13117	T9K3-Sp6.2
35	56.6	5.9	959	29	CNS00655	CNS00655	AL062806
36	56.4	5.8	871	13	BM262815	BM262815	BM262815
37	56	5.8	888	28	AZ528430	AZ528430	ENTC024TR
38	55.8	5.8	481	13	BM268570	BM268570	BM268570
39	55.8	5.8	1101	29	CNS0039R	CNS0039R	Drosophila
40	55.4	5.7	1152	29	BZ696911	BZ696911	SP_Ba009
41	55.2	5.7	644	13	BM136514	BM136514	BM136514
42	55.2	5.7	916	28	AZ671886	AZ671886	ENTLRA4TR
43	55	5.7	742	13	BX462820	BX462820	BX462820
44	54.8	5.7	452	13	BM138725	BM138725	BM138725
45	54.8	5.7	770	12	BM169947	BM169947	EST572470

ALIGNMENTS

RESULT 1
BH161951/c 868 bp DNA linear GSS 24-SEP-2001
LOCUS ENTBR12TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.
ACCESSION BH161951
VERSION BH161951.1 GI:15735389
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun

OY 518 AAAAGCCAGCAGACAGAGCTGATTTGCTCAATTAGCCAAAGATATTCACACTGATG 577
 DB 309 AAGAAAGATGATGAAGAGAGATGATGAAGAGATGATGAAGAGAGATGATG 250
 OY 578 AAAAACAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATG 637
 DB 249 ATGAAGACAGCAGATATTAATTAATTAATTAATTAATTAATTAATTAATTA 190
 OY 638 AGCAAGTCAAAAAGCCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 697
 DB 189 AAGAAAGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 130
 OY 698 CAATGCGACACAGCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 757
 DB 129 ATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70
 OY 758 AAAAATCATCTAATTTGATGCTCAAGAAATTAATAA 799
 DB 69 AAGATGATGAAGATGATGAAGAGAGATGATGAAGATGATGAAGATGATGAAG 28

RESULT 3
 BHI60272 931 bp DNA linear GSS 24-SEP-2001
 LOCUS ENT0V49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION BHI60272
 VERSION BHI60272.1 GI:15733710
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 931)
 AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 8
 High quality sequence stop: 677.
 Location/Qualifiers

FEATURES
 Source
 1..931
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pROSI, Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + 1 method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 402 a 51 c 279 g 199 t

BASE COUNT
 ORIGIN

Query Match 7.4%; Score 71.6; DB 28; Length 931;
 Best Local Similarity 45.2%; Pred. No. 0.068;
 Matches 263; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

OY 218 AAGTTTGAAGAAACATATGCTCAAGCTTGATGATGAAGAGCTTATGATGATG 277
 DB 96 ATGATGATGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATG 155
 OY 278 CCGAAGAAAAAAGACATATGAGCAAACTACCAACGCTGCTTTCACAGCAGGATG 337
 DB 156 AAGAAAGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 215
 OY 338 CTCTTGAAGACAGTAACTCAATTCGTACAGTAATTAAGTGAAGTGGCAGTTAGA 397
 DB 216 ATGATGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 275
 OY 398 AGTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 457
 DB 276 AAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 335
 OY 458 CAGATGTAACGGCTCAATTCATCCGCTTAATTAATGAAGATGAAGCCAAAGAGTTCTG 517
 DB 336 AAGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 395
 OY 518 AAAAAGCCAGCAGAGAGCTGCTGATTTGCTCAATTAAGCCAAAGATATTCACACTGATG 577
 DB 396 AAGAAAGAGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 455
 OY 578 AAAAACAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 637
 DB 456 ATGAAGACAGCAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 515
 OY 638 AGCAAGTCAAAAAGCCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 697
 DB 516 AAGAAAGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 575
 OY 698 CAATGCGACACAGCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 757
 DB 576 ATGATGAAGAGATGATGAAGAGAGATGATGAAGAGAGATGATGAAGAGATGATG 635
 OY 758 AAAAATCATCTAATTTGATGCTCAAGAAATTAATAA 799
 DB 636 AAGATGATGAAGATGATGAAGAGAGATGATGAAGAGATGATGAAGATGATGAAG 677

RESULT 4
 A2531291/c 877 bp DNA linear GSS 03-NOV-2000
 LOCUS ENT0304TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2531291
 VERSION A2531291.1 GI:11085838
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 JOURNAL Unpublished
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 22

ACCESSION A2676392
 VERSION A2676392.1 GI:11813538
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 REFERENCE Eukaryota; Entamoebidae; Entamoeba.
 AUTHORS 1 (bases 1 to 938)
 TITLE Loftus, B., Van Aken, S. and Fraser, C.
 JOURNAL Determination of clone end sequences from Entamoeba histolytica
 COMMENT HMI:IMSS sheared DNA library
 Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 34
 High quality sequence stop: 809.
 Location/Qualifiers
 1..938
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing. projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."
 BASE COUNT 132 a 201 c 107 g 498 t
 ORIGIN
 Query Match 6.88; Score: 66; DB 28; Length 938;
 Best Local Similarity 46.7%; Pred. No. 0.53;
 Matches 210; Conservative 0; Mismatches 240; Indels 0; Gaps 0;
 353 AACCTGAATTCGTACAGTAATTTAGTTGGCAGTTAAGAGCTAGCAGAACTG 412
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db AACCTGAAGCAAAAGCTAAAGAGAGCTAGCAAAAGCTAAAGAGAGCTGAACAA 733
 413 AATTGACAGTAGAGCTATAGAAAGCCCTTGTAGTACACTCCAGATGTAAAGCTC 472
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db AACCTGAAGCAAAAGCTAAAGAGAGCTAAAGAGAGCTGAAGCAAAAGCAAGAG 673
 473 AAATCATCGCTTTAATATGAAGATAGAGCCCAAGAGTTCTCGAAAAAGCCAGG 532
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db AACCTGAAGCAAAAGCTAAAGAGAGCTGAAGCAAAAGCAAGAGCTGAACAA 613
 533 AAGCTGCTGATTTTGCATTTAGCCAAAGATATTCATGATGAGAAAAACAAAGAA 592
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db AACCTGAAGCAAAAGCTAAAGAGAGCTAAAGAGAGCTGAAGCAAAAGCTGAAG 553
 593 ATGGTGAGAAATATCTTGTATTCCTTCAACAGAGTACCTGAGCAAGTCAAAAAG 652
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db AACCTGAAGCAAAAGCTAAAGAGAGCTAAAGAGAGCTAAAGAGAGTAAAGAG 493
 653 CCGCTTTCGCTTATGATGTGATGCTTCTGTATGATTAACAGCAAGTGCACAG 712
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 492 AAATAAAGAGAGCTTAAGAGAGAGCTTAAGAGAGCTTAAGAGAGCTTAAGAGAG 433
 Qy 713 CCTACAGTAGCCCAATTTACATTTGTAAGACTCCTAAGAAACAGAAAAATCTTAATA 772
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 432 AAATAAAGAGAGAGCTTAAGAGAGAGCTTAAGAGAGCTTAAGAGAGCTTAAGAGAG 373
 Qy 773 TTGATGACTCAAGAGAGAGAGCTTAAGAGAGCTTAAGAGAGCTTAAGAGAGCTG 802
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 372 AAATAAAGAGAGAGCTTAAGAGAGAGCTTAAGAGAGCTTAAGAGAGCTG 343
 RESULT 10
 A2530768 890 bp DNA linear GSS 03-NOV-2000
 LOCUS ENTB54YF Entamoeba histolytica sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2530768 GI:11084894
 VERSION A2530768
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 890)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library
 COMMENT Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 20
 High quality sequence stop: 822.
 Location/Qualifiers
 1..890
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing. projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."
 BASE COUNT 375 a 61 c 225 g 229 t
 ORIGIN
 Query Match 6.7%; Score 64.8; DB 28; Length 890;
 Best Local Similarity 46.8%; Pred. No. 0.82;
 Matches 204; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
 218 AAGTTTGAAGAAATATGAGCTGAGCTGATTAAGAGCTGATGATGATGATGATG 277
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 448 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
 278 CCGAAGATG 337
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

508 AACAGATGATGATGCAAGATGATGATGAAGTATGATGAAGATGATGAAGATGATG 567

0Y 338 CTCTTGAACACGCTAAAGCTCAATTGCTACAGTAAATTAGTTGAGTGGCAGTTAAGA 397

Db 568 ATGAAGATGATGATGAAGATGATGATGAAGATGATGAAGATGATGAAGATGATGA 627

0Y 398 AGGTAGCAGAAAGCGATGATGACAGATGAAGCGCTTAAAGAAAGCCCTTGATGAGTAC 457

Db 628 ATGAAGCGATGATGAAGAGCAGATGATAGCGATGATGAAGACGATGATGAAGACGATG 687

0Y 458 CAGATGTAAAGGGCTCAATCATCCGTTAATAATGAAGATTAAGGCCAAAGAGTTCTCG 517

Db 688 ATGATGAAGATGAAGATGAAGACGATGATGATGAAGACGATGAAGACGATGATGATG 747

0Y 518 AAAAAAGCAAAGCGAAGGTGCTGATTTGCTCAATTAGCCAAAGATTAATTCAACTGATG 577

Db 748 AAGTGTGAAGCAGATGATGAAGACGATGATGATGAAGACGATGATGAAGACGATG 807

0Y 578 AAAAAACAAAAGAAATGCTGGGAATTAATTCCTTGATTCGCTTCAACGAAGTACTCG 637

Db 808 ATGATGATGATGAAGAAAGAGAAAGACGCTTATGAAGACATCAGAGAACATTCG 867

0Y 638 AGCAAGTCAAAAAAGC 653

Db 868 GAGACGTTATGAAGAC 883

RESULT	11
BH149983	
LOCUS	
DEFINITION	BH149983 976 bp DNA linear GSS 27-AUG-2001
ACCESSION	ENTQD93JTF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
VERSION	BH149983
KEYWORDS	BH149983.1 GI:15311165
SOURCE	GSS.
ORGANISM	Entamoeba histolytica
REFERENCE	Entamoeba histolytica
AUTHORS	Eukaryota; Entameobidae; Entamoeba.
TITLE	1 (bases 1 to 976)
JOURNAL	Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C.
COMMENT	Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
UNPUBLISHED	
Contact:	Brendan J Lofthus
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel.: 301 838 0208
	Fax: 301 838 3543
	Email: bjl@fusedtigr.org
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
	Seq primer: M13-Forward
	Class: shotgun
	High quality sequence start: 38
	High quality sequence stop: 740.
FEATURES	Location/Qualifiers
SOURCE	1..976

FEATURES	SOURCE	Location/Qualifiers
1.	976	
		/organism="Entamoeba histolytica"
		/mol_type="genomic DNA"
		/strain="HMI:IMSS"
		/db_xref="taxon:5759"
		/clone_id="Entamoeba histolytica Sheared DNA"
		/note="Vector: pHOSt1. Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
		Genomic DNA isolated from broth cultures of <i>E. histolytica</i> using a method described by Clark and Diamond (Clark, C.G., and Diamond, I.S. (1993) <i>Entamoeba histolytica</i> : a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for

Query Match	6.7%:	Score 64.6:	DB 28:	Length 976:
Best Local Similarity	45.7%:	Pred. No. 0.87:	269:	Indels 0:
Matches 226:	Conservative	0:	Mismatches	269:
BASE COUNT	415 a	63 c	280 g	218 t
ORIGIN	Whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).			

Query Match	Best Local Similarity	5.7%;	Score 64.6;	DB 28;	Length 976;
Matches	226;	Conservative	0;	Mismatches 269;	Indels 0;
					Gaps 0;
QY	194	TCCTGTTAAATATGACCATCCAAAAGTTTTGTAAGAAAACATATGCGTCAGAGCTTGATG	253		
Db	95	TATCTTAACGTGTCATGCCGAGACAGAGTTTGTAACTCGATGATGATTTGTAATTTATA	154		
QY	254	ATTAAGAGGTTGATGATCTATTTCGCCAGAGAAAAAACAATATGCGGAAAACATCCAC	313		
Db	155	ATGAAGATTTTGATTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	214		
QY	314	GTCGCTTGTCACAGCAGGATGATGACTCTGTGAACACGTAAAGCTCAATTCGTACAACTA	373		
Db	215	ATGATGATGATTTGGAATTTAGAACACGACGAGAGATGATGATGATGATGATGATGATG	274		
QY	374	AATTAGTTGACTTGGCAGTTAAGAGGTAGCAGAGCTGAATTGACAGATGAACCTATA	433		
Db	275	AAGAAGAAAGAGATGAAGAAGAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG	334		
QY	434	AGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	493		
Db	335	ACGATGATGAAGAAAGACGATGATGAAGAGCAGATGAAGAAAGACGATGATGAAGAG	394		
QY	494	AAGATTAAGGGCCAAAGAAAGTTCTCGAAAAGCCAGAGCGTGTCTATTTTCTTCANT	553		
Db	395	ACGATGATGAAGAAAGACGATGATGAAGAGCAGATGAAGAAAGACGATGATGAAGAG	454		
QY	554	TAGCCAAAGATTAATTCACCTGATGAAAAAACAAAAAGAAATGTTGGAGAAATTAACCTTG	613		
Db	455	ACGATGATGAAGAAAGACGATGATGAAGAGCAGATGATGAAGAAAGATGATGATGAAGATG	514		
QY	614	ATTTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAGCCGCTTTGCTTTAAGATGTGG	673		
Db	515	ATGATGATGAAGATGATGAATTTGATTTAGAAAGCAGATGATGAAGATGATGATGAAGAG	574		
QY	674	ATGGATTTTCGATG 688			
Db	575	ATGATGATGAAGATG 589			

RESULT	12	
LOCUS	AZ669474	
DEFINITION	880 bp DNA linear GSS 14-DEC-2000	
ACCESSION	ENTRIY88TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.	
VERSION	AZ669474	
KEYWORDS	AZ669474.1	GI:11806620
SOURCE	GSS.	
ORGANISM	Entamoeba histolytica	
REFERENCE	Entamoeba histolytica	
AUTHORS	Eukaryota: Entamoebidae; Entamoeba.	
TITLE	1 (bases 1 to 880)	
	Loftus, B., Van Aken, S. and Fraser, C.	
	Determination of clone end sequences from Entamoeba histolytica	
	HM1:IMSS sheared DNA library	
JOURNAL	Unpublished	
COMMENT	Contact: Brendan J Loftus	
	Department of Eukaryotic Genomics	
	The Institute for Genomic Research	
	9712 Medical Center Dr., Rockville, MD 20850, USA	
	Tel: 301 838 0208	
	Fax: 301 838 3543	
	Email: b1o@fuser.tigr.org	
	Clones are derived from the Entamoeba histolytica HM1:IMSS sheared	
	DNA library	
	Seq primer: M13-Reverse	

[illegible][illegible]

```
Db      464 AGATGATGACGTAGATTAATGATGACGAAAGATGACCGTAGAT--GATGATGACGAGA 407
QY      393 TAAGAAGGTAGCAGAAAGCTGAATTGACATGAAAGCCTTAAGAAAGCCTTGATGAGTA 452
Db      406 TGATGACGAAAGATGATGATGACGAAATGATGACGATGATGATGACGAAAGATGA 347
QY      453 CACTCCAGATGTAACGGCTCAATCATCCGCTTATTAATGAAGATAAGGCCAAAGAGT 512
Db      346 CGAAGATGATGAAGAAGATGATGACCTAGATGATGACGATGATGATGACGAAAGATGA 287
QY      513 TCTCGAAAAAGCCAAAGGCGCAAGAGTCTGATTTGCTCAATTAGCCAAAGATATTCAAC 572
Db      286 TGACGAAAGATGATGACGTAGATGATGACGAAATGATGACGAAAGATGATGACGAAAG 227
QY      573 TGATGAAAAACAAAAAATGCTGAGAAATTAACCTTGATTTCTGCTCAACAGAGT 632
Db      226 TGATGACGAAAGATGACGAAAGATGATGACGACGATGATGATGATGATGATGATGATGA 167
QY      633 ACCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATGATGATGATGATGATGATGATGAT 692
Db      166 TGACAAAGATGACGAAAGATGCTGACGAAGATGATGATGATGATGATGATGATGATGAT 107
QY      693 T 693
Db      106 T 106
```

Search completed: September 4, 2003, 21:02:05
Job time : 2074 secs

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

Fri Sep 5 09:24:49 2003

us-10-049-473a-1.rnmp

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:26:32 ; Search time 2926 Seconds

(without alignments)
10818.092 Million cell updates/sec

Title: US-10-049-473a-1

Perfect score: 966

Sequence: 1 agtaacattatctcaaggy.....gcagtagtactcaacagca 966

Scoring table: IDENTITY_NTIC
Gapop 10.0 , Gapext 1.0

Searched: 32997241 seqs, 16383922548 residues

Total number of hits satisfying chosen parameters: 65994482

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.old:*
3: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US098D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US099E_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US099F_COMB.seq:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	966	100.0	966	US-10-049-473a-1	Sequence 1, App1
2	966	100.0	47	US-10-158-844-108	Sequence 108, App
3	965.2	99.9	56	US-60-029-960-148	Sequence 148, App
4	961.2	99.5	60	US-60-068-175-584	Sequence 584, App

Graser
10/049473 Page 1
Seq. 105 1st 2nd
Interf

```

5 939 97.2 942 33 US-09-769-787-179 Sequence 179, App
6 934.2 96.9 942 25 US-09-583-110-2539 Sequence 2539, App
7 868.2 89.9 888 16 US-09-107-433-1524 Sequence 1524, App
8 807.8 83.6 11591 60 US-60-061-998-559 Sequence 559, App
9 552.4 57.2 640 14 US-08-961-083-33 Sequence 33, App
10 552.4 57.2 640 33 US-09-765-271-33 Sequence 33, App
11 552.4 57.2 640 33 US-09-765-272-33 Sequence 33, App
12 552.4 57.2 640 33 US-09-765-272-33 Sequence 33, App
13 265.6 27.5 451 13 US-08-832-030-202 Sequence 202, App
14 265.6 27.5 451 13 US-08-832-030-202 Sequence 202, App
15 265.6 27.5 451 13 US-08-832-030-202 Sequence 202, App
16 216.2 22.4 1110 48 US-09-134-000C-879 Sequence 219, App
17 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 219, App
18 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 219, App
19 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 219, App
20 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 219, App
21 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 219, App
22 211.6 21.9 1035 60 US-60-068-217-711 Sequence 711, App
23 209.8 21.6 2102 59 US-60-068-217-711 Sequence 711, App
24 208.8 21.6 2102 59 US-60-068-217-711 Sequence 711, App
25 208.8 21.6 2102 59 US-60-068-217-711 Sequence 711, App
26 207.2 21.4 7169 60 US-60-068-217-711 Sequence 711, App
27 207.2 21.4 7169 60 US-60-068-217-711 Sequence 711, App
28 206.4 21.4 7169 60 US-60-068-217-711 Sequence 711, App
29 203.2 21.0 970 52 US-10-417-884-1082 Sequence 1082, App
30 151.2 15.7 909 52 US-10-417-884-1082 Sequence 1082, App
31 149.6 15.5 15081 60 US-60-068-186-732 Sequence 732, App
32 149.6 15.5 15081 60 US-60-068-186-732 Sequence 732, App
33 142.4 14.7 1599 58 US-60-068-186-732 Sequence 732, App
34 142.4 14.7 1599 58 US-60-068-186-732 Sequence 732, App
35 79.2 8.2 17475 29 US-09-663-779-1357 Sequence 1357, App
36 76.6 7.9 43980 51 US-10-398-221-2058 Sequence 2058, App
37 76.6 7.9 43980 51 US-10-398-221-2058 Sequence 2058, App
38 73.4 7.6 340 58 US-60-045-653-1195 Sequence 1195, App
39 73.4 7.6 340 58 US-60-045-653-1195 Sequence 1195, App
40 71.8 7.4 1720 9 US-10-398-221-3334 Sequence 3334, App
41 59.8 6.2 7218 9 US-08-468-194-14 Sequence 14, App
42 59.2 6.1 1163 51 US-10-398-221-1250 Sequence 1250, App
43 54 5.6 2205 49 US-10-266-090-32333 Sequence 32333, App
44 53.8 5.6 2135 49 US-10-282-122A-16758 Sequence 16758, App
45 53.4 5.5 1206 49 US-10-266-090-24388 Sequence 24388, App

```

ALIGNMENTS

```

RESULT 1
US-10-049-473a-1
: Sequence 1, Application US/10049473a
: GENERAL INFORMATION:
: APPLICANT: de Groot, Ronald
: TITLE OF INVENTION: Pneumococcal Vaccines
: FILE REFERENCE: Docket 294-120 PCT/US
: CURRENT APPLICATION NUMBER: US/10/049, 473a
: PCT FILING DATE: 2002-02-12
: PRIOR APPLICATION NUMBER: PCT/NL00/00569
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: EP 99202640.1
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 966
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
US-10-049-473a-1

```

```

Query Match 100.0%; Score 966; DB 44; Length 966;
Best Local Similarity 100.0%; Pred. No. 8.9e-186;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 AGTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTTATGCGAGGTGCATCACA 60

```

```

Db
1 AGTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTTATGCGAGGTGCATCACA 60
61 CTATTATCAGTAGCACTTATAGCAGCTTGTTCGAAAGGTCAGAAAGGTGCAGACCTATC 120
61 CTATTATCAGTAGCACTTATAGCAGCTTGTTCGAAAGGTCAGAAAGGTGCAGACCTATC 120
121 AGCATGAAAGGAGTGCATTCACGAACATCAATTTATGAGCAAGTGAAGCAACCTT 180
121 AGCATGAAAGGAGTGCATTCACGAACATCAATTTATGAGCAAGTGAAGCAACCTT 180
181 TCAGCCCAACAGCTTGTTCATTAATATGACCATCCAAAGTTTGTGAAGCAATATGCG 240
181 TCAGCCCAACAGCTTGTTCATTAATATGACCATCCAAAGTTTGTGAAGCAATATGCG 240
241 TCAGAGCTTATGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATGATG 300
241 TCAGAGCTTATGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATGATG 300
301 GAAACTACCAAGCTGCTTGTTCACAGAGGATGATGATGATGATGATGATGATGATG 360
301 GAAACTACCAAGCTGCTTGTTCACAGAGGATGATGATGATGATGATGATGATGATG 360
361 ATTGCTACCAAGTAAATTAATGATGATGATGATGATGATGATGATGATGATGATG 420
361 ATTGCTACCAAGTAAATTAATGATGATGATGATGATGATGATGATGATGATGATG 420
421 GATGAGCCCTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 480
421 GATGAGCCCTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 480
481 GATGAGCCCTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 540
481 GATGAGCCCTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 540
541 GATTTGCTCAATTAAGCAAGTAAATTAATGATGATGATGATGATGATGATGATGATG 600
541 GATTTGCTCAATTAAGCAAGTAAATTAATGATGATGATGATGATGATGATGATGATG 600
601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
661 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
661 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
721 AGCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
721 AGCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
781 TACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
781 TACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
841 CAAGCATTAATGAGAAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATG 900
841 CAAGCATTAATGAGAAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATG 900
901 CAAATATATCTTACCAATATATGCGTGTGATGATGATGATGATGATGATGATGATG 960
901 CAAATATATCTTACCAATATATGCGTGTGATGATGATGATGATGATGATGATGATG 960
961 AAGCAA 966
961 AAGCAA 966
Db

```

```

RESULT 2
US-10-158-844-108
: Sequence 108, Application US/10158844
: GENERAL INFORMATION:
: APPLICANT: Kunsch et al.

```



```

: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 391
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: CD-R
: COMPUTER: Dell Latitude Pentium 3
: OPERATING SYSTEM: Windows 98
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/158,844
: FILING DATE: 03-Jun-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/961,527
: FILING DATE: 1997-10-30
: APPLICATION NUMBER: US 60/029,960
: FILING DATE: 1996-10-31
: ATTORNEY/AGENT INFORMATION:
: NAME: Hyman, Mark J.
: REGISTRATION NUMBER: 46,789
: REFERENCE/DOCKET NUMBER: PB340PID1
: INFORMATION FOR SEQ ID NO: 108:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11309 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-10-158-844-108

Query Match      100.0%; Score 966; DB 47; Length 11309;
Best Local Similarity 100.0%; Pred. No. 1.6e-187;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

: 1 AGTAACCTTATCTCAAGAGATGACATGCAAGAAAAAATATTTGGCAGGTGCCATCACA 60
: 632 AGTAACCTTATCTCAAGAGATGACATGCAAGAAAAAATATTTGGCAGGTGCCATCACA 7691
: 61 CTATATCAAGTACCACTTATGACAGTGTGTCGAAGGTCAGAGAAGTGCAGACTTATC 120
: 7692 CTATATCAAGTACCACTTATGACAGTGTGTCGAAGGTCAGAGAAGTGCAGACTTATC 7751
: 121 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTT 180
: 7752 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTT 7811
: 181 TCAGCCCAACAATCTGTTGAATATGACATCCAAAAAGTTTTCAGAAAAACAATATGGC 240
: 7812 TCAGCCCAACAATCTGTTGAATATGACATCCAAAAAGTTTTCAGAAAAACAATATGGC 7871
: 241 TCAGACTTGATGATTAAGAGTGTGATGATCTATTGCGCAAGAAAAAACAATATGGC 300
: 7872 TCAGACTTGATGATTAAGAGTGTGATGATCTATTGCGCAAGAAAAAACAATATGGC 7931
: 301 GAAACCTACCAAGCTGCTTGTGACACAGAGTATGACTCTTGAACACCTAAAGCTCAA 360
: 7932 GAAACCTACCAAGCTGCTTGTGACACAGAGTATGACTCTTGAACACCTAAAGCTCAA 7991
: 361 ATTCGTACAGTAATTAATAGTTGAGTGGCAGTTAAGAGGTACAGAAAGCTGAATTGACA 420
: 7992 ATTCGTACAGTAATTAATAGTTGAGTGGCAGTTAAGAGGTACAGAAAGCTGAATTGACA 8051
: 421 GATGAAGCCTATTAAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGATGAT 480
: 8052 GATGAAGCCTATTAAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGATGAT 8111
: 481 CGCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

```

```

: 8112 CGCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8171
: 541 GATTTCCTCAATATACCAAGATATATTCACATGATGATGATGATGATGATGATGATGATGAT 600
: 8172 GATTTCCTCAATATACCAAGATATATTCACATGATGATGATGATGATGATGATGATGATGAT 8231
: 601 GAAATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
: 8232 GAAATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8291
: 661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
: 8292 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8351
: 721 AGCCATATATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
: 8352 AGCCATATATACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8411
: 781 TACAAAGAAAAAATTAATACTGTTATCTGATGATGATGATGATGATGATGATGATGATGATGAT 8471
: 8412 TACAAAGAAAAAATTAATACTGTTATCTGATGATGATGATGATGATGATGATGATGATGATGAT 8531
: 841 CAAAGCATATATCGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
: 8472 CAAAGCATATATCGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9591
: 901 CAAATATCTTTTCCCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
: 8532 CAAATATCTTTTCCCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8591
: 961 AACGAA 966
: 8592 AACGAA 8597

RESULT 3
US-60-029-960-148/c
: Sequence 148, Application US/60029960
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 1649
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/60/029,960
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340PP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 148:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11167 base pairs
: TYPE: nucleic acid

```

STRANDEDNESS: double
TOPOLOGY: Linear
US-60-029-960-148

Query Match
Best Local Similarity 99.9%; Score 965.2; DB 56; Length 11167;
Matches 964; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 AGTAACTTATCTCAAAAGAGTAGACATGAGAAAAATTTATGGCAGGTGCCATCACA 60
3693 AGTAACTTATCTCAAAAGAGTAGACATGAGAAAAATTTATGGCAGGTGCCATCACA 3634
61 CTATTATCACTAGCACTTTAGCAAGCTTTGCCAAAGGTCAGAGAGTGCAGCTTATC 120
3633 CTATTATCACTAGCACTTTAGCAAGCTTTGCCAAAGGTCAGAGAGTGCAGCTTATC 3574
121 AGCATGAAAGGGAGTGCATTACAGAACATCAATTTATGAGCAAGTGAAGAACCCCT 180
3573 AGCATGAAAGGGAGTGCATTACAGAACATCAATTTATGAGCAAGTGAAGAACCCCT 3514
181 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 240
3513 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 3454
241 TCAGAGCTTATGATTAAGAGGTTGATGATTAATTTGCCAGAAAAACAATATGCG 300
3453 TCAGAGCTTATGATTAAGAGGTTGATGATTAATTTGCCAGAAAAACAATATGCG 3394
301 GAAACTACCAAGCTGCTTGTCAACAAGATATGATCTTGTGAAACAGTAAAGCTCAA 360
3393 GAAACTACCAAGCTGCTTGTCAACAAGATATGATCTTGTGAAACAGTAAAGCTCAA 3334
361 ATTCTGTAAGTAATTAAGTGAAGTGCAGTTAGAGGTAGCAGAGTGAATGACA 420
3333 ATTCTGTAAGTAATTAAGTGAAGTGCAGTTAGAGGTAGCAGAGTGAATGACA 3274
421 GATGAAGCTTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 480
3273 GATGAAGCTTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 3214
481 CGTCTTAATTAATGAGTAAGGCAAGAAAGTCTCGAAAAAGCCAAAGGAGAGTGC 540
3213 CGTCTTAATTAATGAGTAAGGCAAGAAAGTCTCGAAAAAGCCAAAGGAGAGTGC 3154
541 GATTTGCTCAATTAAGCAAGTAATTAATCAAGTGAAGAAAAACAAGAAATGCTGCA 600
3153 GATTTGCTCAATTAAGCAAGTAATTAATCAAGTGAAGAAAAACAAGAAATGCTGCA 3094
601 GAAATTAACCTTGTGATCTGCTCAACAAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 660
3093 GAAATTAACCTTGTGATCTGCTCAACAAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 3034
661 GCTTTAGAGTGTGATGCTGCTCAACAAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 720
3033 GCTTTAGAGTGTGATGCTGCTCAACAAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 2974
721 AGCCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
2973 AGCCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2914
781 TACAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
2913 TACAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2854
841 CAAAGCATTAATCGAAAAAGATTAAGCAAGCAATATCAAGGTAAAGCAAGCCCTTC 900
2853 CAAAGCATTAATCGAAAAAGATTAAGCAAGCAATATCAAGGTAAAGCAAGCCCTTC 2794
901 CAAAGCATTAATCGAAAAAGATTAAGCAAGCAATATCAAGGTAAAGCAAGCCCTTC 960
2793 CAAAGCATTAATCGAAAAAGATTAAGCAAGCAATATCAAGGTAAAGCAAGCCCTTC 2734
961 AACGAA 966

Db 2733 AACGAA 2728

RESULT 4

US-60-068-175-584

Sequence 584, Application US/60068175

GENERAL INFORMATION:

APPLICANT: Lagace, Robert E.

APPLICANT: Corley, Neil C.

APPLICANT: Russo, Frank D.

APPLICANT: Hann, Amy L.

APPLICANT: Finney, Gregory L.

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STREPTOCOCCUS PNEUMONIAE

NUMBER OF INVENTIONS: GENOME, FRAGMENTS THEREOF, AND USES THEREOF

NUMBER OF SEQUENCES: 1175

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/60/068, 175

FILING DATE: HEREWITH

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PM-0009-2 P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-416

INFORMATION FOR SEQ ID NO: 584:

SEQUENCE CHARACTERISTICS:

LENGTH: 12356 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: genomic DNA

IMMEDIATE SOURCE:

CLONE: SPN2586

US-60-068-175-584

Query Match
Best Local Similarity 99.5%; Score 961.2; DB 60; Length 12356;
Matches 963; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AGTAACTTATCTCAAAAGAGTAGACATGAGAAAAATTTATGGCAGGTGCCATCACA 60
8278 AGTAACTTATCTCAAAAGAGTAGACATGAGAAAAATTTATGGCAGGTGCCATCACA 8337
61 CTATTATCACTAGCACTTTAGCAAGCTTTGCCAAAGGTCAGAGAGTGCAGCTTATC 120
8338 CTATTATCACTAGCACTTTAGCAAGCTTTGCCAAAGGTCAGAGAGTGCAGCTTATC 8397
121 AGCATGAAAGGGAGTGCATTACAGAACATCAATTTATGAGCAAGTGAAGAACCCCT 180
8398 AGCATGAAAGGGAGTGCATTACAGAACATCAATTTATGAGCAAGTGAAGAACCCCT 8457
181 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 240
8458 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 8517
241 TCAGAGCTTATGATTAAGAGGTTGATGATTAATTTGCCAGAAAAACAATATGCG 300

```

RESULT 5
US-09-769-787-179
: Sequence 179: Application US/09769787
: GENERAL INFORMATION:
: APPLICANT: Microbial Technics Limited
: APPLICANT: Gilbert, Christophe FG
: APPLICANT: Hansbro, Philip M
: TITLE OF INVENTION: Proteins
: FILE REFERENCE: PMC/P2129MO
: CURRENT APPLICATION NUMBER: US/09/769,787
: CURRENT FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: GB 9816337.1
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: US 60/125164
: PRIOR FILING DATE: 1999-03-19
: NUMBER OF SEQ ID NOS: 388
: SOFTWARE: Seqalin Ver. 2.1
: SEQ ID NO 179
: LENGTH: 942
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: US-09-769-787-179

```

QY	28	ATGAAGAAAAATTAATTGGCAGAGTGCCTCATCACTAATTATGATGCAACTTATACAGT	87
Db	1	ATGAAGAAAAATTAATTGGCAGAGTGCCTCATCACTAATTATGATGCAACTTATACAGT	60
QY	88	TGTTCGAAAGGGTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGATGTCTATACAGAA	147
Db	61	TGTTCGAAAGGGTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGATGTCTATACAGAA	120
QY	148	CATCAATTTTATGAGCAAGTAAAAAGCAACCCCTTCAGGCCAACAAGTCTTGTAATATG	207
Db	121	CATCAATTTTATGAGCAAGTAAAAAGCAACCCCTTCAGGCCAACAAGTCTTGTAATATG	180
QY	208	ACCATCCAAAAAGTTTGTGAAAAAACAAATNTGGCTCAGAGCTTGATGATAAAGAGGTGAT	267
Db	181	ACCATCCAAAAAGTTTGTGAAAAAACAAATNTGGCTCAGAGCTTGATGATAAAGAGGTGAT	240
QY	268	GATACTATTTGCCGAGAAAAAAAACAATATGCGCAAAACTACCAACGTGTCTTGACAA	327
Db	241	GATACTATTTGCCGAGAAAAAAAACAATATGCGCGAAAACTACCAACGTGTCTTGACAA	300
QY	328	GCAGGTATGACTCTTGAAACAGTAACTCAATTCGTACAAAGTAATAGTTAGTTG	387
Db	301	GCAGGTATGACTCTTGAAACAGTAACTCAATTCGTACAAAGTAATAGTTAGTTG	360
QY	388	GCAGTTAAGAAGTAGCAGAAAGCTGAAATGACAGATGAAGCCTATAGAAGCCCTTGAT	447
Db	361	GCAGTTAAGAAGTAGCAGAAAGCTGAAATGACAGATGAAGCCTATAGAAGCCCTTGAT	420
QY	448	GAGTACACTCCAGATGTAAAGGCTCAATCATCGTCTTAATATGAAGATTAAGCCAAA	507
Db	421	GAGTACACTCCAGATGTAAAGGCTCAATCATCGTCTTAATATGAAGATTAAGCCAAA	480
QY	508	GAACTTTCGAAAAAGCCAAAGGCAAGAGTGTGATTTGTCATATTGGCCAAAGATAT	567
Db	481	GAACTTTCGAAAAAGCCAAAGGCAAGAGTGTGATTTGTCATATTGGCCAAAGATAT	540
QY	568	TCAACTGATGAAAAAACAAAAAGAAATGTGAGAGAAATTAACCTTGGATTGCTTCAACA	627
Db	541	TCAACTGATGAAAAAACAAAAAGAAATGTGAGAGAAATTAACCTTGGATTGCTTCAACA	600
QY	628	GAACTACTGAGCAAGTCAAAAAAGCCGCTTTCGTTTATGATGTGATGTGTTCTGAT	687
Db	601	GAACTACTGAGCAAGTCAAAAAAGCCGCTTTCGTTTATGATGTGATGTGTTCTGAT	660
QY	688	GTGATTTACAGAACTGGGCACAACAGCCCTACAGTACCAATATTAATTGTAAACCTCACT	747
Db	661	GTGATTTACAGAACTGGGCACAACAGCCCTACAGTACCAATATTAATTGTAAACCTCACT	720
QY	748	AAGAAAAACAGAAAAATCATCTTAATTTATGACTCAAAAGAAAAATTTAAAAACCTTTATC	807
Db	721	AAGAAAAACAGAAAAATCATCTTAATTTATGACTCAAAAGAAAAATTTAAAAACCTTTATC	780
QY	808	TTGACTCAAAAAACAAATGATTCAACATTTGTTCAACAGCATTTATCGAAAAAGATTGCA	867
Db	781	TTGACTCAAAAAACAAATGATTCAACATTTGTTCAACAGCATTTATCGAAAAAGATTGCA	840
QY	868	GCAGCCAATATCAAGGTTTAAGGACCAGCCTTCCAAAATATCTTTATACCAATATATCGGT	927
Db	841	GCAGCCAATATCAAGGTTTAAGGACCAGCCTTCCAAAATATCTTTATACCAATATATCGGT	900
QY	928	GGTGAGATTCAGAGTCAAGAGTGTACATTAACGAA	966
Db	901	GGTGAGATTCAGAGTCAAGAGTGTACATTAACGAA	939

RESULT 6
US-09-583-110-2539
; Sequence 2539, Application US/09583110


```

OY 83 CACCTTGTGGAAGG-GTCAGAAAGTGCAGACCTTATCAGCATGAAGGGAGTGCATT 141
DB 1 CACCTTGTGGAAGG-GTCAGAAAGTGCAGACCTTATCAGCATGAAGGGAGTGCATT 60
OY 142 ACAGAACATCAATTTTATGAGCAAGTGAAGAACACCCCTTACAGCCCAAGCTTGT 201
DB 61 ACAGAACATCAATTTTATGAGCAAGTGAAGAACACCCCTTACAGCCCAAGCTTGT 120
OY 202 AATATGACCATCCAAAAGTTTGAAGAAACATATGAGCTCAGAGCTTATGATTAAG 261
DB 121 AATATGACCATCCAAAAGTTTGAAGAAACATATGAGCTCAGAGCTTATGATTAAG 180
OY 262 GTTATGATATCTATGCGGAGAAAGAAACATATGAGCGAAAGTACCAAGCTTGT 321
DB 181 GTTATGATATCTATGCGGAGAAAGAAACATATGAGCGAAAGTACCAAGCTTGT 240
OY 322 TCACAGCAGTATGAGTCTTGAAGAACCTTAAGCTCAATTTGTATCAATTAAT 381
DB 241 TCACAGCAGTATGAGTCTTGAAGAACCTTAAGCTCAATTTGTATCAATTAAT 300
OY 382 GACTTGGCAGTATGAAGAGTACGAGAGTGAATGACAGTGAAGCTTAAAGAAC 441
DB 301 GACTTGGCAGTATGAAGAGTACGAGAGTGAATGACAGTGAAGCTTAAAGAAC 360
OY 442 TTTGATGATACACTCCAGATGTAAGGCTCAATATCCGCTTATTAATGAAGAT 501
DB 361 TTTGATGATACACTCCAGATGTAAGGCTCAATATCCGCTTATTAATGAAGAT 420
OY 502 GCCAAGAAAGTCTCGAAAAGGCCAGAGGCTGATTTTGTCAATTAAGCCAA 561
DB 421 GCCAAGAAAGTCTCGAAAAGGCCAGAGGCTGATTTTGTCAATTAAGCCAA 480
OY 562 GATATTCACACTGATGAAAAACAAAAGAAATGTTGAGAAATTAACCTTGT 621
DB 481 GATATTCACACTGATGAAAAACAAAAGAAATGTTGAGAAATTAACCTTGT 540
OY 622 TCACAGAAAGTACCTGAGCAAGTCAAAAAGCCGCTTTCGCTTGAAGTGT 681
DB 541 TCACAGAAAGTACCTGAGCAAGTCAAAAAGCCGCTTTCGCTTGAAGTGT 600
OY 682 TCGATGATGATTAACAGCAAGTGCACAGACCTTACAGTGAAGTATTAAT 741
DB 601 TCGATGATGATTAACAGCAAGTGCACAGACCTTACAGTGAAGTATTAAT 660
OY 742 CTCACATAGAAAGAGAAATATCATTAATTTGATGATGATGATGATGAT 801
DB 661 CTCACATAGAAAGAGAAATATCATTAATTTGATGATGATGATGATGAT 720
OY 802 GTTATCTTGAAGTCAAAAAGAAATGATTCACATTTGTTCAAGCATTA 861
DB 721 GTTATCTTGAAGTCAAAAAGAAATGATTCACATTTGTTCAAGCATTA 780
OY 862 TTTCAAGCAGCCTATATCAAGGTTAAGAGCAGCCTTCCAAATATCTT 921
DB 781 TTTCAAGCAGCCTATATCAAGGTTAAGAGCAGCCTTCCAAATATCTT 840
OY 922 ATCGGTGATGATTCACAGTCAAGCAGTATGATGATGATGATGATGAT 966
DB 841 ATCGGTGATGATTCACAGTCAAGCAGTATGATGATGATGATGATGAT 885

```

RESULT 8

US-60-061-998-559

```

; Sequence 559, Application US/60061998
; GENERAL INFORMATION:

```

```

; APPLICANT: LAGACE, ROBERT E.
; APPLICANT: CORLEY, NEIL C.
; APPLICANT: RUSSO, FRANK D.
; APPLICANT: HANN, AMY L.
; APPLICANT: HEATH, JOE D.
; APPLICANT: FINNEY, GREGORY L.

```

```

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STREPTOCOCCUS PNEUMONIAE
; GENOME, FRAGMENTS THEREOF, AND USES THEREOF

```

```

; NUMBER OF SEQUENCES: 797
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/061,998
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PM-0006-2P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEO ID NO: 559:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: SP1c564
; US-60-061-998-559

```

```

Query Match 83.68; Score 807.8; DB 60; Length 11591;
Best Local Similarity 98.98; Pred. No. 4.5e-155;
Matches 824; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

```

```

OY 1 AGTAACACTTATTCGAAAGAGTACATGAGAAAGAAATTAATGCGAGGTGCATAC 60
DB 6276 AGTAACACTTATTCGAAAGAGTACATGAGAAAGAAATTAATGCGAGGTGCATAC 6335
OY 61 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 120
DB 6336 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 6395
OY 121 AGCATGAAGGGGATGTCATTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 180
DB 6396 AGCATGAAGGGGATGTCATTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 6455
OY 181 TCGACCCCAAGAGCTGTTAAATATGACATCCAAAAGTTTGAAGAAACAATATG 240
DB 6456 TCGACCCCAAGAGCTGTTAAATATGACATCCAAAAGTTTGAAGAAACAATATG 6515
OY 241 TCAGAGCTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGAT 300
DB 6516 TCAGAGCTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGAT 6575
OY 301 GAAACTACCAAGCTGCTTGTCAAGACAGATGACTCTTGAACAGCTAAAGCTCAA 360
DB 6576 GAAACTACCAAGCTGCTTGTCAAGACAGATGACTCTTGAACAGCTAAAGCTCAA 6635
OY 361 ATTGCTACAGTAAATAGTTGAGTGGCAGTTAAGAGGATGAGAGGATGATGACA 420
DB 6636 ATTGCTACAGTAAATAGTTGAGTGGCAGTTAAGAGGATGAGAGGATGATGACA 6695
OY 421 GATCAAGCTATTAAGAAAGCTTTGATGATGATGATGATGATGATGATGATGAT 478
DB 6696 GATCAAGCTATTAAGAAAGCTTTGATGATGATGATGATGATGATGATGATGAT 6755
OY 479 TCGCTTTAATTAAGAAAGTGAAGCCAAAGAGTTCGAAAAAGCCAAAGCAAGGTG 538
DB 6755 TCGCTTTAATTAAGAAAGTGAAGCCAAAGAGTTCGAAAAAGCCAAAGCAAGGTG 538

```

Db 6756 TCCGCTTAATATGATAGATAGCCCAAGAGTTCTCGAAAAAGCCAGCAGAGGTG 6815
OY 539 CTGATTTTGCATTTAGCCAAAGATATATTCATCTGATGAAAAAACAAGAAATGCTG 598
Db 6816 CTGATTTTGCATTTAGCCAAAGATATATTCATCTGATGAAAAAACAAGAAATGCTG 6875
OY 599 GAGAAATTAAGCTTGTCTGCTTCACAGAGTACCTGACCAAGTCAAAAAAGCCGCTT 658
Db 6876 GAGAAATTAAGCTTGTCTGCTTCACAGAGTACCTGACCAAGTCAAAAAAGCCGCTT 6935
OY 659 TCGCTTATGATGTGAGTGTCTTCTGATGTGATTAAGCAAGTGGCAGACAGCTTCA 718
Db 6936 TCGCTTATGATGTGAGTGTCTTCTGATGTGATTAAGCAAGTGGCAGACAGCTTCA 6995
OY 719 GTACCAATATATTAATGTAATTAAGTCACTAAGAAAAAATCATCTATATGATG 778
Db 6996 GTACCAATATATTAATGTAATTAAGTCACTAAGAAAAAATCATCTATATGATG 7055
OY 779 ACTCAAGAAAAATTAATTAAGTCACTAAGAAAAAATCATCTATATGATG 831
Db 7056 ACTCAAGAAAAATTAATTAAGTCACTAAGAAAAAATCATCTATATGATG 7108

RESULT 9

US-08-961-083-33
Sequence 33, Application US/08961083

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 640 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 33:

US-08-961-083-33

Query Match 57.2%; Score 552.4; DB 14; Length 640;

Best Local Similarity 94.7%; Pired. No. 5.4e-103;

Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

OY 90 TTCGAAAGGTCAGAGTGTACAGCTTATCAGATGAAAGGGGATGATTAACAGAA 149

Db 1 TTCGAAAGGTCAGAGTGTACAGCTTATCAGATGAAAGGGGATGATTAACAGAA 60

OY 150 TCAATTTATGAGCAAGTGAAGAACCTTCAGCCCAACAGCTTGTAAATATGAC 209
Db 61 TCAATTTATGAGCAAGTGAAGAACCTTCAGCCCAACAGCTTGTAAATATGAC 120
OY 210 CATCAAAAAATTTTGAACAAATATGCTCAGAGCTTGATGATAAGAGTTGATGA 269
Db 121 CATCAAAAAATTTTGAACAAATATGCTCAGAGCTTGATGATAAGAGTTGATGA 180
OY 270 TACTATTGCGGCAAAAAAACAATATGGGCAAAATACCAAGCTGTTGTCAGAAC 329
Db 181 TACTATTGCGGCAAAAAAACAATATGGGCAAAATACCAAGCTGTTGTCAGAAC 240
OY 330 AGTATGACTCTTGAACACAGTAAAGCTCAAAATCTGACAGTAAATTAAGTGTGGC 389
Db 241 AGTATGACTCTTGAACACAGTAAAGCTCAAAATCTGACAGTAAATTAAGTGTGGC 300
OY 390 AGTTAAGAGGTACAGAGCTGAATGACAGATGAAGCCTATTAAGAACCTTTGATGA 449
Db 301 AGTTAAGAGGTACAGAGCTGAATGACAGATGAAGCCTATTAAGAACCTTTGATGA 360
OY 450 GTACACTCCAGATGTACAGGCTCAATCCGCTTAATTAATGAATGAAGCCAAAGA 509
Db 361 GTACACTCCAGATGTACAGGCTCAATCCGCTTAATTAATGAATGAAGCCAAAGA 420
OY 510 AGTTCTGAAAAAGCCAGGACAGAGGTGCTGATTTGCTCAATTAAGCCAAAGTAAATTC 569
Db 421 AGTTCTGAAAAAGCCAGGACAGAGGTGCTGATTTGCTCAATTAAGCCAAAGTAAATTC 480
OY 570 AACTGATGAAGAAAAAAGAAATGTTGAGAAATTAAGTAAATTAAGTAAATTAAGTAAATTC 629
Db 481 AACTGATGAAGAAAAAAGAAATGTTGAGAAATTAAGTAAATTAAGTAAATTAAGTAAATTC 540
OY 630 AGTACCTG--AGCAAGTCAAAAAAGCCGCTTTCGCTTATGATGATGATGATGATGAT 687
Db 541 AGTACCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATGATGATGATGATGATGATGAT 600
OY 688 G-----TGATTACAGCAACTGCGACACA 710
Db 601 GGAATGGAATTAACGCAACTGCGGACACA 628

RESULT 10

US-09-765-271-33
Sequence 33, Application US/09765271

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,271

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/536,784

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 33:
 SEQUENCE CHARACTERISTICS:

LENGTH: 640 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 33:
 US-09-765-272-33

Query Match 57.2%; Score 552.4; DB 33; Length 640;
 Best Local Similarity 94.7%; Pred. No. 5.4e-103;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

```

QY 90 TTCGAAGGCTCAGAGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGACA 149
   |||||
DB 1 TTCGAAGGCTCAGAGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGACA 60
   |||||

QY 150 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 209
   |||||
DB 61 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 120
   |||||

QY 210 CATCCAAAAGTTTGGAAAACAAATATGCTCAGAGCTTGATATAAGAGTTGATGA 269
   |||||
DB 121 CATCCAAAAGTTTGGAAAACAAATATGCTCAGAGCTTGATATAAGAGTTGATGA 180
   |||||

QY 270 TACTATTGCCGAAGAAAACAAATATGCGGAAAACCTACCAAGTCTTGTACAAAGC 329
   |||||
DB 181 TACTATTGCCGAAGAAAACAAATATGCGGAAAACCTACCAAGTCTTGTACAAAGC 240
   |||||

QY 330 AGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAAATAGTTAGTTGGC 389
   |||||
DB 241 AGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAAATAGTTAGTTGGC 300
   |||||

QY 390 AGTTAAGAGTGAAGCAAGCTGAATTCAGATGAAGCCTATAGAAACCTTTGATGA 449
   |||||
DB 301 AGTTAAGAGTGAAGCAAGCTGAATTCAGATGAAGCCTATAGAAACCTTTGATGA 360
   |||||

QY 450 GTACACTCCAGATGTAAAGGCTCAATTCATCCCTTTAATATGAAGATAAGGCCAAAGA 509
   |||||
DB 361 GTACACTCCAGATGTAAAGGCTCAATTCATCCCTTTAATATGAAGATAAGGCCAAAGA 420
   |||||

QY 510 AGTTCTCGAAAAGCCAGAGCAGAGTGTGATTTTGTCAATTAGCCAAAGATAATTC 569
   |||||
DB 421 AGTTCTCGAAAAGCCAGAGCAGAGTGTGATTTTGTCAATTAGCCAAAGATAATTC 480
   |||||

QY 570 AACTGTAGAAAACAAAAGAAAATGTGTGAGAAATTAACCTTGATTTGCTTCAACAGA 629
   |||||
DB 481 AACTGTAGAAAACAAAAGAAAATGTGTGAGAAATTAACCTTGATTTGCTTCAACAGA 540
   |||||

QY 630 AGTACCTG--AGCAAGTCAAAAAGCCGCTTTCGCTTTAGATGTGATGTGTTCTGAT 687
   |||||
DB 541 AGTACCTGAGCAAGTCAAAAAGCCGCTTTCGCTTTAGATGTGATGTGTTCTGAT 600
   |||||

QY 688 G-----TGATTACAGCACTGCACACA 710
   |||||
DB 601 GATGTGATTAACGCAACTGGGGCACA 628
   |||||

```

RESULT 11

US-09-765-272-33

Sequence 33, Application us/09765272

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 640 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 33:
 US-09-765-272-33

Query Match 57.2%; Score 552.4; DB 33; Length 640;
 Best Local Similarity 94.7%; Pred. No. 5.4e-103;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

```

QY 90 TTCGAAGGCTCAGAGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGACA 149
   |||||
DB 1 TTCGAAGGCTCAGAGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGACA 60
   |||||

QY 150 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 209
   |||||
DB 61 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 120
   |||||

QY 210 CATCCAAAAGTTTGGAAAACAAATATGCTCAGAGCTTGATATAAGAGTTGATGA 269
   |||||
DB 121 CATCCAAAAGTTTGGAAAACAAATATGCTCAGAGCTTGATATAAGAGTTGATGA 180
   |||||

QY 270 TACTATTGCCGAAGAAAACAAATATGCGGAAAACCTACCAAGTCTTGTACAAAGC 329
   |||||
DB 181 TACTATTGCCGAAGAAAACAAATATGCGGAAAACCTACCAAGTCTTGTACAAAGC 240
   |||||

QY 330 AGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAAATAGTTAGTTGGC 389
   |||||
DB 241 AGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAAATAGTTAGTTGGC 300
   |||||

QY 390 AGTTAAGAGTGAAGCAAGCTGAATTCAGATGAAGCCTATAGAAACCTTTGATGA 449
   |||||
DB 301 AGTTAAGAGTGAAGCAAGCTGAATTCAGATGAAGCCTATAGAAACCTTTGATGA 360
   |||||

QY 450 GTACACTCCAGATGTAAAGGCTCAATTCATCCCTTTAATATGAAGATAAGGCCAAAGA 509
   |||||
DB 361 GTACACTCCAGATGTAAAGGCTCAATTCATCCCTTTAATATGAAGATAAGGCCAAAGA 420
   |||||

QY 510 AGTTCTCGAAAAGCCAGAGCAGAGTGTGATTTTGTCAATTAGCCAAAGATAATTC 569
   |||||
DB 421 AGTTCTCGAAAAGCCAGAGCAGAGTGTGATTTTGTCAATTAGCCAAAGATAATTC 480
   |||||

QY 570 AACTGTAGAAAACAAAAGAAAATGTGTGAGAAATTAACCTTGATTTGCTTCAACAGA 629
   |||||
DB 481 AACTGTAGAAAACAAAAGAAAATGTGTGAGAAATTAACCTTGATTTGCTTCAACAGA 540
   |||||

QY 630 AGTACCTG--AGCAAGTCAAAAAGCCGCTTTCGCTTTAGATGTGATGTGTTCTGAT 687
   |||||
DB 541 AGTACCTGAGCAAGTCAAAAAGCCGCTTTCGCTTTAGATGTGATGTGTTCTGAT 600
   |||||

```

202-050-750 00 00

```

QY      688 G-----TGATTACAGCAACTGGCAGACA 710
        | | | | | | | | | | | | | | | | | |
Db      601 GGATGTGGATTACAGCAACTGGGCACACA 628

RESULT 12
US-09-765-272a-33
/ Sequence 33, Application US/09765272A
/ GENERAL INFORMATION:
/   APPLICANT: Choi et. al.
/   TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
/   Vaccines
/   NUMBER OF SEQUENCES: 454
/   CORRESPONDENCE ADDRESSES:
/     ADDRESSEE: Human Genome Sciences, Inc.
/     STREET: 9410 Key West Avenue
/     CITY: Rockville
/     STATE: Maryland
/     COUNTRY: USA
/     ZIP: 20850
/
/ COMPUTER READABLE FORM:
/   MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
/   COMPUTER: Dell Latitude C610
/   OPERATING SYSTEM: Windows 2000
/   SOFTWARE: ASCII Text
/
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/09/765,272A
/   FILING DATE: 22-Jan-2001
/   CLASSIFICATION: <Unknown>
/   PRIOR APPLICATION DATA:
/     APPLICATION NUMBER: 08/961,083
/     FILING DATE: OCT-30-1997
/
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Lin J. Hymel
/   REGISTRATION NUMBER: 45,414
/   REFERENCE/DOCKET NUMBER: PB340P2C2
/   TELECOMMUNICATION INFORMATION:
/     TELEPHONE: (301) 610-5790
/     TELEFAX: (301) 309-8439
/
/ INFORMATION FOR SEQ ID NO: 33:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 640 base pairs
/     TYPE: nucleic acid
/     STRANDEDNESS: double
/     TOPOLOGY: linear
/   SEQUENCE DESCRIPTION: SEQ ID NO: 33:
/
US-09-765-272a-33

Query Match      57.2%; Score 552.4; DB 33; Length 640;
Best Local Similarity 94.7%; Pred. No. 5.4e-103;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2

          90  TTGAAAGGGCACAAGGTCCAGACCTTATCAGCATGAAGGAGTGCATTACAGAAC 149
          | | | | | | | | | | | | | | | | | |
Ddb      1  TTCAAAGGGTCACAAGGTCCAGACCTTATCAGCATGAAGGAGTGCATTACAGAAC 60

          150 TCAATTTTATGAGCAAGTGAAGCAACCTTCAGGCCCAACAGTCTGTTAAATATGAC 209
          | | | | | | | | | | | | | | | | | |
Ddb      61 TCAATTTTATGAGCAAGTGAAGCAACCTTCAGGCCCAACAGTCTGTTAAATATGAC 120

          210 CATCCAAAAAGTTTTTAAATAACAATATGAGCTCAGAGCTTGATGATTAAGAGGTGATGA 269
          | | | | | | | | | | | | | | | | | |
Ddb      121 CATCCAAAAAGTTTTTGAATAACAATATGAGCTCAGAGCTTGATGATTAAGAGGTGATGA 180

          270 TACTATGCGCGAAGAAAAAACAATATGAGCGAAACTACCAACGTGTCTTGACACAGC 329
          | | | | | | | | | | | | | | | | | |
Ddb      181 TACTATGCGCGAAGAAAAAACAATATGAGCGAAACTACCAACGTGTCTTGACACAGC 240

          330 AGGTATGACTCTTGAAAACACGTAAGCTCAATTCGTACAGTAATATGATGAGTGGC 389
          | | | | | | | | | | | | | | | | | |
Ddb      241 AGGTATGACTCTTGAAAACACGTAAGCTCAATTCGTACAGTAATATGATGAGTGGC 300

```


Query Match 27.5%; Score 265.6; DB 13; Length 451;
Best Local Similarity 97.9%; Pred. No. 3.2e-44;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTATCTCAAGAGTAGACATGAAGAAAAAATTAATGGCAGGTGCATCACA 60
|||||
Db 167 AGTAACACTATCTCAAGAGTAGACATGAAGAAAAAATTAATGGCAGGTGCATCACA 226
|||||
QY 61 CTAATTACATACCACTTTAGCAGCTTTGTCGAAAGGTCAGAGGTGACCTTATC 120
|||||
Db 227 CTAATTACATACCACTTTAGCAGCTTTGTCGAAAGGTCAGAGGTGACCTTATC 286
|||||
QY 121 AGCATGAAGGGATGTCATTACGAACATCAATTTATGACGAAGTGAAGAACACCT 180
|||||
Db 287 AGCATGAAGGGATGTCATTACGAACATCAATTTATGACGAAGTGAAGAACACCT 346
|||||
QY 181 TCAGCCCAACAGCTTTGTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 239
|||||
Db 347 TCAGCCCAACAGCTTTGTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 406
|||||
QY 240 CTCAGAGCTTGATGATTAAGAGGTGATGATTAATTTGCCGAGA 284
|||||
Db 407 CTCAGAGCTTGATGATTAAGAGGTGATGATTAATTTGCCGAGA 451
|||||

RESULT 14

PCT-US97-07950-219

Sequence 219, Application PC/TUS9707950

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: Novel Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/07950
FILING DATE: 14-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 219:
SEQUENCE CHARACTERISTICS:
LENGTH: 791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
PCT-US97-07950-219

Query Match 27.5%; Score 265.6; DB 1; Length 791;
Best Local Similarity 97.9%; Pred. No. 3.7e-44;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTATCTCAAGAGTAGACATGAAGAAAAAATTAATGGCAGGTGCATCACA 60
|||||
Db 507 AGTAACTATCTCAAGAGTAGACATGAAGAAAAAATTAATGGCAGGTGCATCACA 566
|||||
QY 61 CTAATTACATACCACTTTAGCAGCTTTGTCGAAAGGTCAGAGGTGACCTTATC 120
|||||
Db 567 CTAATTACATACCACTTTAGCAGCTTTGTCGAAAGGTCAGAGGTGACCTTATC 626
|||||
QY 121 AGCATGAAGGGATGTCATTACGAACATCAATTTATGACGAAGTGAAGAACACCT 180
|||||
Db 627 AGCATGAAGGGATGTCATTACGAACATCAATTTATGACGAAGTGAAGAACACCT 686
|||||
QY 181 TCAGCCCAACAGCTTTGTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 239
|||||
Db 687 TCAGCCCAACAGCTTTGTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 746
|||||
QY 240 CTCAGAGCTTGATGATTAAGAGGTGATGATTAATTTGCCGAGA 284
|||||
Db 747 CTCAGAGCTTGATGATTAAGAGGTGATGATTAATTTGCCGAGA 791
|||||

RESULT 15

PCT-US97-07950-219

Sequence 219, Application PC/TUS9707950

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: Novel Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/07950
FILING DATE: 14-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 219:
SEQUENCE CHARACTERISTICS:
LENGTH: 791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
PCT-US97-07950-219

Query Match

27.5%; Score 265.6; DB 2; Length 791;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 20:27:27 ; Search time 187 Seconds
(without alignments)
5693.830 Million cell updates/sec

Title: US-10-049-473a-1
Perfect score: 966
Sequence: 1 agtaacctatctcaag.....gcagtagcatcaacga 966

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 907674 seqs, 55110846 residues

Total number of hits satisfying chosen parameters: 1815348

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database : Pending_Patents_NA_New:*
1: /cgn2_6/pdata/2/pna/PC7_NEW_COMB.seq:*
2: /cgn2_6/pdata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/pdata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/pdata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/pdata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/pdata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/pdata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	934.2	96.7	942	6	US-10-640-833-2539	Sequence 2539, Ap
2	868.2	89.9	888	6	US-10-617-320-1524	Sequence 1524, Ap
3	52	5.4	1107	6	US-10-603-113-5692	Sequence 5692, Ap
4	51.4	5.3	672	6	US-10-603-113-7512	Sequence 7512, Ap
5	46.2	4.8	1263	6	US-10-602-494-313	Sequence 1842, Ap
6	45.2	4.7	3095	6	US-10-602-494-313	Sequence 313, App
7	44.6	4.6	17703	6	US-10-257-1668-34	Sequence 34, App
8	44.4	4.6	1287	6	US-10-603-113-3698	Sequence 3698, Ap
9	44.2	4.6	342	6	US-10-603-113-5931	Sequence 5931, Ap
10	43.8	4.5	96596	1	PCR-US02-38582-100	Sequence 100, App
11	43.4	4.5	3231	6	US-10-257-1668-164	Sequence 164, App
12	43.2	4.5	653	6	US-10-626-717-3404	Sequence 3404, Ap
13	43	4.5	2250	6	US-10-602-494-302	Sequence 302, App
14	43	4.5	5898	6	US-10-640-833-1168	Sequence 1168, Ap
15	43	4.5	5919	6	US-10-617-320-648	Sequence 648, App
16	43	4.5	50000	5	US-09-662-2548-23	Sequence 23, App
17	42.2	4.4	891	6	US-10-603-113-11247	Sequence 11247, A
18	42.2	4.4	1835	6	US-10-612-783-2093	Sequence 2093, Ap
19	42.2	4.4	2754	6	US-10-603-113-4857	Sequence 4857, Ap
20	42.2	4.4	19920	6	US-10-257-1668-5	Sequence 5, Appl
21	42	4.3	17220	6	US-10-602-494-327	Sequence 327, App
22	41.8	4.3	7892	6	US-10-257-1668-138	Sequence 138, App
23	41.6	4.3	2082	6	US-10-603-113-4064	Sequence 4064, Ap
24	41.2	4.3	40178	6	US-10-600-009-467	Sequence 467, App
25	41.2	4.3	40178	6	US-10-600-009-468	Sequence 468, App
26	41.2	4.3	50423	7	US-60-485-450-12101	Sequence 12101, A

27	41	4.2	453	6	US-10-640-833-1702	Sequence 1702, Ap
28	41	4.2	477	6	US-10-617-320-1055	Sequence 1055, Ap
29	41	4.2	1367	6	US-10-302-172-553	Sequence 553, App
30	40.8	4.2	1785	6	US-10-603-113-928	Sequence 928, App
31	40.8	4.2	1791310	5	US-09-947-914-46	Sequence 46, Appl
32	40.6	4.2	5271	6	US-10-450-826-125	Sequence 125, App
33	40.6	4.2	22268	6	US-10-450-826-126	Sequence 126, Appl
34	40.4	4.2	3463	6	US-10-299-636-93	Sequence 93, Appl
35	40.4	4.2	17220	6	US-10-602-494-185	Sequence 185, App
36	40.2	4.2	3561	6	US-10-603-113-2080	Sequence 2080, Ap
37	40.2	4.2	11473	6	US-10-257-1668-111	Sequence 111, App
38	40	4.1	2377	6	US-10-602-494-231	Sequence 231, App
39	40	4.1	2469	6	US-10-603-113-4746	Sequence 4746, Ap
40	39.8	4.1	2229	6	US-10-602-494-240	Sequence 240, App
41	39.8	4.1	2382	6	US-10-602-494-325	Sequence 325, App
42	39.6	4.1	1269	6	US-10-603-113-1601	Sequence 1601, Ap
43	39.6	4.1	7441	6	US-10-257-1668-140	Sequence 140, App
44	39.6	4.1	7615	6	US-10-257-1668-53	Sequence 53, Appl
45	39.6	4.1	171596	7	US-60-487-610-19488	Sequence 19488, A

ALIGNMENTS

RESULT 1
US-10-640-833-2539
; Sequence 2539, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qilong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PAT03-15
; CURRENT APPLICATION NUMBER: US/10/640, 833
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583, 110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107, 433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085, 131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051, 553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 2539
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-2539

Query Match 96.7%; Score 934.2; DB 6; Length 942;
Best Local Similarity 99.7%; Pred. No. 3.7e-247;
Matches 936; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	28	ATGACAAATATATATGTCAGTGCATCACAATATATGCTAGCACTTAGAGCT	87
DB	1	ATGACAAATATATATGTCAGTGCATCACAATATATGCTAGCACTTAGAGCT	60
QY	88	TGTTCAAGGCGCAGACGTCAGCTTATCAGCATGAAGGCGATCTTACAGAA	147
DB	61	TGTTCAAGGCGCAGACGTCAGCTTATCAGCATGAAGGCGATCTTACAGAA	120
QY	148	CATCAATTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAACTCTGTTAATATG	207
DB	121	CATCAATTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAACTCTGTTAATATG	180
QY	208	ACCATCAAAAGCTTTTGAACAAATATGCTGACACTGATCAATAAAGCTTAT	267
DB	181	ACCATCAAAAGCTTTTGAACAAATATGCTGACACTGATCAATAAAGCTTAT	240

```

.OY 268 GATCTATTCGCGAAGAAAAAACAATATGCGGAAAACTACCAAGCTGTCTGTCAAA 327
    |||||||
Db 241 GATCTATTCGCGAAGAAAAAACAATATGCGGAAAACTACCAAGCTGTCTGTCAAA 300
.OY 328 GCAGTATGACTCTGTAAGAACGTAATGCTCAATTTGTCACAGTAATAGTTGAGTTG 387
    |||||||
Db 301 GCAGTATGACTCTGTAAGAACGTAATGCTCAATTTGTCACAGTAATAGTTGAGTTG 360
.OY 388 GCAGTATGAGAGTAGCAGAGGAGTGAATGACATGACCTATATGAAGAGCTTTGAT 447
    |||||||
Db 361 GCAGTATGAGAGTAGCAGAGGAGTGAATGACATGACCTATATGAAGAGCTTTGAT 420
.OY 448 GAGACACTCCAGATGTAAAGCGCTCAATATCCGCTTATATATGAAGTAAGGCCAAA 507
    |||||||
Db 421 GAGACACTCCAGATGTAAAGCGCTCAATATCCGCTTATATATGAAGTAAGGCCAAA 480
.OY 508 GAAGTTCTCGAAAAAGCGAAGGAGAGTGTGATTTGCTCAATTAAGCAAAAGTAAT 567
    |||||||
Db 481 GAAGTTCTCGAAAAAGCGAAGGAGAGTGTGATTTGCTCAATTAAGCAAAAGTAAT 540
.OY 568 TCAACTGATGAAAAAACAAGAAATGCTGAGAAATTTACCTTTGATTCGCTTCACA 627
    |||||||
Db 541 TCAACTGATGAAAAAACAAGAAATGCTGAGAAATTTACCTTTGATTCGCTTCACA 600
.OY 628 GAAGTACCTGAGAGTCAAAAAAGCCGCTTTCGCTTATGATGTGATGCTTTGAT 687
    |||||||
Db 601 GAAGTACCTGAGAGTCAAAAAAGCCGCTTTCGCTTATGATGTGATGCTTTGAT 660
.OY 688 GTGATTTACAGCACTGCGACACAGCCTACAGTACCAATATATGATTTGAAAAAGCAGCT 747
    |||||||
Db 661 GTGATTTACAGCACTGCGACACAGCCTACAGTACCAATATATGATTTGAAAAAGCAGCT 720
.OY 748 AGAAGAACGAAAAATCATCTATATTTGATGACTACAAAGAAAAATTAAGCTGTTATC 807
    |||||||
Db 721 AGAAGAACGAAAAATCATCTATATTTGATGACTACAAAGAAAAATTAAGCTGTTATC 780
.OY 808 TTGACTCAAAAAACAATATGATTCATTTGTCAAAGATTTACGAAAAAGATTTGCAA 867
    |||||||
Db 781 TTGACTCAAAAAACAATATGATTCATTTGTCAAAGATTTACGAAAAAGATTTGCAA 840
.OY 868 GGAAGCAATATCAAGTTAAAGACCAAGCCTTCACAAATATCTTTACCAATATATCGGT 927
    |||||||
Db 841 GGAAGCAATATCAAGTTAAAGACCAAGCCTTCACAAATATCTTTACCAATATATCGGT 900
.OY 928 GGTGAGATTCAAGCTCAAGCAGTAGTACATCAAAAGAA 966
    |||||||
Db 901 GGTGAGATTCAAGCTCAAGCAGTAGTACATCAAAAGAA 939

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariulello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO.1524:
SEQUENCE CHARACTERISTICS:
LENGTH: 888 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc.feature
LOCATION: (8) LOCATION 1..888
SEQUENCE DESCRIPTION: SEQ ID NO: 1524:
US-10-617-320-1524

Query Match      89.9%; Score 868.2; DB 6; Length 888;
Best Local Similarity 99.5%; Pred. No. 5.1e-229;
Matches 881; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

.OY 83 CAGCTTTGTCGAAGG-GTCAGAGGTGCGACCTTATCAGCATGAAGGGAGTGCATT 141
    |||||||
Db 1 CAGCTTTGTCGAAGG-GTCAGAGGTGCGACCTTATCAGCATGAAGGGAGTGCATT 60
.OY 142 ACAGAACATCAATTTTATGAGCAAGTGAAGGCAACCCCTCAGCCCAAGTCTGTGTA 201
    |||||||
Db 61 ACAGAACATCAATTTTATGAGCAAGTGAAGGCAACCCCTCAGCCCAAGTCTGTGTA 120
.OY 202 AATATGACCATCCAAAAAGTTTGAAGAACATATGCTCAGAGCTGATGATTAAGAG 261
    |||||||
Db 121 AATATGACCATCCAAAAAGTTTGAAGAACATATGCTCAGAGCTGATGATTAAGAG 180
.OY 262 GTTGATGATACATTTGGCGAAGAAAAAACAATATGCGGAAAACTACCAAGCTGTG 321
    |||||||
Db 181 GTTGATGATACATTTGGCGAAGAAAAAACAATATGCGGAAAACTACCAAGCTGTG 240
.OY 322 TCACAGCAGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAATAGTT 381
    |||||||
Db 241 TCACAGCAGGTATGACTCTTGAAGACAGTAAAGCTCAATTCGTACAAAGTAATAGTT 300
.OY 382 GAGTGGCAGTTAAGAGAGTGCAGAGCTAATTTGACAGATGAAGCCTATTAAGAAAGCC 441
    |||||||
Db 301 GAGTGGCAGTTAAGAGAGTGCAGAGCTAATTTGACAGATGAAGCCTATTAAGAAAGCC 360
.OY 442 TTTGATGATACATCCAGATGTAAAGGCTCAATTCGCTTAATTAAGAGATTAAG 501
    |||||||
Db 361 TTTGATGATACATCCAGATGTAAAGGCTCAATTCGCTTAATTAAGAGATTAAG 420
.OY 502 GCCAAAGAGTTCTCGAAAAAGCCAGAGCAGAGGCTGATTTTGTCAATTAAGCCAAA 561
    |||||||
Db 421 GCCAAAGAGTTCTCGAAAAAGCCAGAGCAGAGGCTGATTTTGTCAATTAAGCCAAA 480
.OY 562 GATATTCACATGATGAAAAAACAAGAAATGCTGAGAAATATACCTTTGATTTGCT 621
    |||||||
Db 481 GATATTCACATGATGAAAAAACAAGAAATGCTGAGAAATATACCTTTGATTTGCT 540
.OY 622 TCAACAGAGTACCTGAGCAAGTCAAAAAAGCCCTTTCGCTTATGATGATGATGCT 681
    |||||||

```

```

RESULT 2
US-10-617-320-1524
; Sequence 1524, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003

```

Db	541	TCACAGAGTACCTGAGCAAGTCATAAAAAAGCCCTTCCGCTTAGATGAGATGGTGT	600
Qy	682	TCGTATGTGATTACAGCAACTGGCAGCACAAAGCCTACAGTAGCCAAATATTACATTGTTAAA	741
Db	601	TCGTATGTGATTACAGCAACTGGCAGCACAAAGCCTACAGTAGCCAAATATTACATTGTTAAA	660
Qy	742	CTCACTAAGAAAACAGAAAAATCATCTAATTTGATGACTCAAGAAGAAAAATTTAAAACT	801
Db	661	CTCACTAAGAAAACAGAAAAATCATCTAATTTGATGACTCAAGAAGAAAAATTTAAAACT	720
Qy	802	GTTATCTTGACTCAAAAAATGATTCAACATTTGTTCAAGCATTATCGAAAAAA	861
Db	721	GTTATCTTGACTCAAAAAATGATTCAACATTTGTTCAAGCATTATCGAAAAAA	780
Qy	862	TTGCACAGCCCAATATCAAGGTTAAGGACCAAGCCTCCAAAATATCTTTACCAATAT	921
Db	781	TTGCACAGCCCAATATCAAGGTTAAGGACCAAGCCTCCAAAATATCTTTACCAATAT	840
Qy	922	ATCGGTGGTGAGATTCAAAGCTCAAGAGTAGTGCATCAAAACGAA	966
Db	841	ATCGGTGGTGAGATTCAAAGCTCAAGAGTAGTGCATCAAAACGAA	885

	RESULT 3	
	US-10-603-113-5692	
	: Sequence 5692, Application US/10603113	
	: GENERAL INFORMATION:	
	: APPLICANT: Kelch Welnsstock et al	
	: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN	
	: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS	
	: FILE REFERENCE: 107196.132	
	: CURRENT APPLICATION NUMBER: US/10/603,113	
	: CURRENT FILING DATE: 2003-06-24	
	: PRIOR APPLICATION NUMBER: US/09/248,796	
	: PRIOR FILING DATE: 1999-02-12	
	: NUMBER OF SEQ ID NOS: 28206	
	: SEQ ID NO 5692	
	: LENGTH: 1107	
	: TYPE: DNA	
	: ORGANISM: Candida albicans	
	US-10-603-113-5692	
	Query March	
	Best Local Similarity 5.4%: Score 52; DB 6; Length 1107;	
	Matches 193; Conservative 0; Mismatches 225; Indels 0; Gaps 0;	
Dc	227 AAAAAAATATGCGCTCAGAGCTTGATGATAAGAAGTTGATGATACATTATGCCGAAGAA 286	
Dc	140 AAGATGAAGTAGAAGAAGAACGATGAGAAAGAACGATGAGAACGAGATGAGATG 199	
Oy	287 AAAACAATATGGCGAAAACCTCACAGCTGTCTGTCAACAAGCAGATGACTTGTAAA 346	
Dc	200 AAGATGACGTTTTTGAATACATGTTGTTGTAATATCCCAAAGCATCATATCACACAA 259	
Oy	347 CACGTAAAGCTCAATTCGTACAGTAATTAATAGTTGAGTTGGCAGTTAAGAAGTAGCAG 406	
Dc	260 CTCGTGATTTACATTAATCTACATGACATGAAGAAGTTATTTGTTGTTAACTGTTCAATCC 319	
Oy	407 AACGTGAATTCACAGATGAACCTATTAAGAAAGCCTTTGATGAGTACCTCCAGATGTAA 466	
Dc	320 CAATTCATTTGCATCGTATATTAATTTGNGCATCCAGCTGATCAAGATGAAAGAAATATG 379	
Oy	467 CGGCTCAATGATCCGCTTTAATATGAAAGTAAAGGCCAAAGAGTTCTGCAAAAAGCCA 526	
Dc	380 ATGATGAGATGAAGATTTACGATGATGATATGATTTTGTCTCCGATGAAAGTGAATCA 439	
Oy	527 AGCGAAGAAGTGTGATTTTGTCTCAATTAGCCAAAGATTAATTCACATGATGAAAAACAA 586	
Dc	440 TTATATGATGCTGCAATTTAGATGATGATATATGACGATGAAGAAGAAAGTGAAGAAGSAA 499	
Oy	587 AAGAAATATGCTGAGAAATTTACCTTTGATTTGCTTCAACGAAGTACCTGACGAATCA 646	
Dc	500 CTCCAAGATTTGAAGAAATTTGTAAGAAAAAAGAAAAAGTTAAAGAAATCACCA 559	

```

RESULT 4
US-10-603-113-7512
: Sequence 7512, Application US/10603113
: GENERAL INFORMATION:
: APPLICANT: Keith Weinstock et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.132
: CURRENT APPLICATION NUMBER: US/10/603.113
: PRIOR FILING DATE: 2003-06-24
: PRIOR APPLICATION NUMBER: US/09/248,796
: PRIOR FILING DATE: 1999-02-12
: NUMBER OF SEQ ID NOS: 28206
: SEQ ID NO 7512
: LENGTH: 672
: TYPE: DNA
: ORGANISM: Candida albicans
US-10-603-113-7512

```

[illegible]

```

RESULT 5
; Sequence 1842, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 1842
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (157),(173)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknd
US-10-603-113-1842

Query Match      4.8%; Score 46.2; DB 6; Length 1263;
Best Local Similarity 47.0%; Pred. No. 0.0068;
Matches 178; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

QY 253 GATTAAGAGGTTGATGATCTATTGCGAAGAAAAAACAATATGCGAAGAACTACCA 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 226 GATGATGAAAAAGGACATGCCAATGGAACAAATGAAATGATCGATATGAAAGT 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 313 CGTGTCTTGTCACAGAGATGACCTGTAAGAACACCTTAAGCTCAATTCGTACAGT 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 286 GAATTAATCATCAGATGGAAGAAATTTGCTGATGAAACCGTTAATGATAAACGTACAGATTA 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 AAATTAAGTGTGAGCAGTTAAGAGTAGCAGAACTGAATTCAGATGAGAGCCTAT 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 GCTAAACAAATATTAGAAATTGGAACAAATGAACATGATTAATGATGAAATTTGAT 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 433 AAGAAACCTTTGATGATGATCACTCCAGATGTAACGGCTTAATTCCTCTTAATTAAT 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 406 GCTAAGATTTAATGATGATGATTAATTAAGCAAGAAATTCACAAAGACATTCGCTGAACCT 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 493 GAAGATTAAGGCCAAGAAAGTTCTC--GAAGAAAGCCAGCAGAGGTGCTGATTTTGGT 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 466 AAGGATACCTTTATTAATTTTTCGGTGATTAATAAATAGATCAATTAATGAGCATAT 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 550 CAATTAGCCAAAGATATTAATCACTGATGAAAAACAAAGAAATGTTGGAGAAATTTACC 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 526 CCATTGATTTAATTAATCACTAAGAAATTTGTTGAAAAATTTAAGTATGACAAATTAAT 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 TTTGATTTCTGCTCAACAG 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 586 TATCATTTTATATATCTG 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

; SEQ ID NO 313
; LENGTH: 3095
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-602-494-313

Query Match      4.7%; Score 45.2; DB 6; Length 3095;
Best Local Similarity 50.5%; Pred. No. 0.018;
Matches 110; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 705 CACCAAGCCTACAGTAGCCAAATATTAATGTAAGTAAGTCACTAAGAAAAACAGAAAAATC 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2084 CAAACAAACAAACAAACCTCCCTTTTACCTCCCAAAATTAACCAAAAAACAAAAA 2025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 765 ATCTAATATTTGATGACATGACAAAGAAAAATTAATACTGTTATCTGACCAAAAAA 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2024 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAACCAAAAAACAAAAA 1965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 825 TGATTCACTTTTGTTCAAAGCATTTATCGAAAAAGAAATTCGAAGCAGCCATATCAAGT 884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1964 CAAAAAAACAAATTTACAAACAAAAAAAAAAAAAAAAAAAAAACACACCAAAATCCAAA 1905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 885 TAAGACCAAGCCTTCACAAATATCTTTACCAATTA 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1904 AATTCACAAACCATACACAAACAAACAAAAAACA 1867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
; Sequence 34, Application US/10257166B
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166B
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 34
; LENGTH: 17703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA
US-10-257-166B-34

Query Match      4.6%; Score 44.6; DB 6; Length 17703;
Best Local Similarity 46.9%; Pred. No. 0.051;
Matches 172; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

QY 558 CAAAGATTAATTCAGTATGAAAAACAAAAAGAAATGGTGAGAAATTTCTTTGATTTC 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15799 CATAAAAAACCACTAATATATATACAAAAATTAATTAATCTACATATCATATTAACCTT 15740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 618 TGCTTCAAGAGAGTACCTGAGCAAGT-CAAAAAAGCGGCTTTCGTTTGATGATGATG 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15739 AAATTAATTAATTAATTAATTTCCACTTAATAAAAAAAAAAACTAACTAAATTAATTA 15680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 677 GTGTTTGTATGTGATTATGAGCACTGCGACAGACAGCTACAGCAATATTAATTAATG 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15679 ACATTAACCAACTATATATACCTACAAAAAACACATCTACCAAAAAAACATATATA 15620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 737 TAAAGTCACATAGAAAAACAGAAAAATCATCTAATTTGATGACTACAAAGAAAAATTA 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

; SEQ ID NO 5931
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-5931

Query Match
  Best Local Similarity 4.6%; Score 44.2; DB 6; Length 342;
  Matches 118; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Db      15619 TAAAAATTAATAAATAAATAATATTCATACAAATAAATAAACAACAA 15560
QY      797  AAACGTGTATCTGACACAAAACAAATGATTCACATCTTGTCAAAGCATTAACGAA 856
Db      15559 TACCTATATTATTAATAAATAAACAACCTTTAAATCAAACTATTAATAAATAAACA 15500
QY      857  AAGAAATTCAGACGACCAATATCAAGTTAAGGACCAAGCCCTTCCAAATATCTTTACCC 916
Db      15499 ATCAATACATTAATAATAAATAAATAATCAATGACAAATAATCCATTAATAAATTCCTACCT 15440
QY      917  AATATAT 923
Db      15439 AATTAAT 15433

```

```

RESULT 8
US-10-603-113-3698
; Sequence 3698, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603.113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248.796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 3698
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-3698

```

```

Query Match
  Best Local Similarity 4.8%; Score 44.4; DB 6; Length 1287;
  Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

```

QY      381  TGAGTTGGCAGTTAAGAGGTAGCAGAACTGATGACGATGAGGAGCCATTAAGAAAGC 440
Db      213  TAATTTAGGAGTATGATGATAGGCTATTAATGATTCAAAGATGAACATTTGAAGAACCA 272
QY      441  CTTTGATGAGTACACTCCAGATGTAAAGCGCTCAATCATCCCTTTAATATGAAGATAA 500
Db      273  CGGTGATGGGATCAATTAAGATAGTGATTAATAGTCAAGATCTTATGACGAGATGA 332
QY      501  GGCCTAAGAAAGTTCTCGAAAAAGCCAAAGCAGAAAGTGTGATTTGCTCAATTAAGCAA 560
Db      333  AGACGATGAGAGTACGACGACGAGATGATGAAGACGAAGAGATGAAGATGAAGTACTA 392
QY      561  AGATTAATTCAGATGAGAAAAAACAAGAAATGGTGAAGAAATTTACCTTGATTTCTGC 620
Db      393  CGACGATCAGCATTTTGAAGAACATTAATGAAGGATTAAGTTGTTGAACCTTAATGAAAA 452
QY      621  TTCAAC 626
Db      453  CTCAC 458

```

```

RESULT 9
US-10-603-113-5931
; Sequence 5931, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603.113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248.796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206

```

```

; SEQ ID NO 5931
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-5931

```

```

Query Match
  Best Local Similarity 4.6%; Score 44.2; DB 6; Length 342;
  Matches 118; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```

```

QY      610  TTGATTTCTGCTTCAACAGAGTACCTGACCAAGTCAAAAAGCCGTTTCCTTAGAT 669
Db      4      TGTATATATATATAAAGCTAAATATTTTCAAGATGATTAATTCATATATCTAGTAAT 63
QY      670  GTGATGCTGTTTCTGATGATTAAGCAACAGCCGACACAGCCATACATACCAATAT 729
Db      64      TTAAATGCTGAATGATGATTAATTCGAGCTTTAAAGATGATGAATGAATTAATAT 123
QY      730  TACATTTGTAACCTCAAGAAACAGAAATATCATTAATTTGATGACTACAGAA 789
Db      124  AAGAAATTTAAACTCCAAATATTTTCAAAATCATATCAAAACCACTACAAAGT 183
QY      790  AAATTAATACTGTATCTTACTCAAAAACAAATGATCAACATTTGTCAAAGCAT 849
Db      184  GATGAATAAATAATTTAATAGCTAATTAAGAAAAAGATCAATTAATTAATTTT 243
QY      850  A 850
Db      244  A 244

```

```

RESULT 10
PCT-US02-38582-100/c
; Sequence 100, Application PC/TUS0238582
; GENERAL INFORMATION:
; APPLICANT: SAGRES DISCOVERY
; APPLICANT: MORRIS, David W.
; APPLICANT: ENGELHARD, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000143
; CURRENT APPLICATION NUMBER: PCT/US02/38582
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/997,722
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 96596
; TYPE: DNA
; ORGANISM: Homo Sapien
PCT-US02-38582-100

```

```

Query Match
  Best Local Similarity 4.5%; Score 43.8; DB 1; Length 96596;
  Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

QY      188  AACAGTCTGTAAATATGACATCCAAAAGTTTGGAAAAACATATGCGTCAGAGC 247
Db      7691  AAAAAATTTCTTAAACATGCTGGAATAAATTTGCCAAGAGATGCTGTATAT 7632
QY      248  TTGATGATTAAGAGGTGATGATACATATGCGCGAAGAAAAAACAATATGCGAAACT 307
Db      7631  CTATGTAATATTAAGTGTGATTAATAATGCAATGAAATATATTTCTTTTGAATAAT 7572
QY      308  ACCAAGCTGCTTGTCAACAGAGTGAAGCTCTTGAACGTAAGCTCAATTCGTA 367
Db      7571  GCAAGCAATATGAGAGAGCAAGTTAATTTCAATAGAACTAACTTTGTTCTG 7512
QY      368  CAAGTAATAGTTGAGTT 386
Db      7511  TATTAATATTTCTTGAGCT 7493

```

RESULT 11

US-10-257-166B-164/c
; Sequence 164, Application US/10257166B
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166B
; PRIOR FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 164
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA
US-10-257-166B-164

Query Match 4.5%; Score 43.4; DB 6; Length 3231;
Best Local Similarity 51.3%; Pred. No. 0.058;
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 716 ACAGTAGCCATATTCATCTTAACCTAGTAAGAAAGCAAAATCATCTATTTG 775
DB 2050 ACAAATACATCTTAACTTAAATCTCATATCAATCAAAAAAATATATATACA 1991
QY 776 ATGACTACAAAGAAAATTAATGCTTATCTGACTCAAAACAAATGATTCACAT 835
DB 1990 TACTCCTTAAATCTATATAATTAATTAACCTTCCACTAAATATACCATACAA 1931
QY 836 TTGTTCAAGCATATGCGAAAGAAATGCAAGACGCAATATCAAGTTAAGACCAAG 895
DB 1930 TATTTATACAAATTTTCAAAAAATACACAAACCCTCAATATCAATTAACCTTAA 1871
QY 896 CCTTCAAAATATCTTT 912
DB 1870 CTATTAATCTAAATTT 1854

RESULT 12

US-10-626-717-3404/c
; Sequence 3404, Application US/10626717
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: De la Pena, Robert C.
; TITLE OF INVENTION: NOCULEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15878)D
; CURRENT APPLICATION NUMBER: US/10/626,717
; PRIOR FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 10/304,123
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: US 09/594,596
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 10952
; SEQ ID NO 3404
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: uc-tsfl9316081c11b1
US-10-626-717-3404

Query Match 4.5%; Score 43.2; DB 6; Length 653;
Best Local Similarity 45.5%; Pred. No. 0.036;
Matches 153; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 559 AAGATATTCACACTGATGAGAAAACAAAGAAATGCTGAGAAATTACCTTGATCT 618
DB 541 AAGATATTAATTAAGAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 482
QY 619 GCTTCAACAGAAAGTACTGAGCAAGTCAAAAAGCCCTTCCCTTGATGATGATGT 678
DB 481 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 422
QY 679 GTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
DB 421 ATATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 362
QY 739 AACTCAGTAAGAAACAGAAAATTCATTAATTAATTAATTAATTAATTAATTAAT 798
DB 361 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 302
QY 799 ACTGTTATCTTGACTCAAAAACAAATGATTCACAAATTTGTTCAAGCATTAATCGA 858
DB 301 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 242
QY 859 GAATTCAGACGACCAATATCAAGTTAAGACCAA 894
DB 241 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 206

RESULT 13

US-10-602-494-302/c
; Sequence 302, Application US/10602494
; GENERAL INFORMATION:
; APPLICANT: Cathy Lofton-Day
; APPLICANT: Andrew Sledziwski
; APPLICANT: Jeff Thomas
; APPLICANT: Robert W. Day
; APPLICANT: Lori Tonnes-Priddy
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of colorectal cell
; FILE REFERENCE: 47675-45
; CURRENT APPLICATION NUMBER: US/10/602,494
; PRIOR FILING DATE: 2003-06-23
; NUMBER OF SEQ ID NOS: 385
; SEQ ID NO 302
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-602-494-302

Query Match 4.5%; Score 43; DB 6; Length 2350;
Best Local Similarity 51.9%; Pred. No. 0.066;
Matches 97; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 738 AACTCAGTAAGAAACAGAAAATTCATTAATTAATTAATTAATTAATTAATTAAT 797
DB 2051 AAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAATTAAT 1992
QY 798 AACTGTTATCTTGACTCAAAAACAAATGATTCACAAATTTGTTCAAGCATTAATCGAA 857
DB 1991 AAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAATTAAT 1932
QY 858 AAGATTCAGACGACCAATATCAAGTTAAGACCAACCTTCCAAATATATCTTACCA 917
DB 1931 ACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1872
QY 918 ATATATC 924
DB 1871 ACAAAC 1865


```

RESULT 14
US-10-640-833-1168
; Sequence 1168, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweat, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PAT#03-15
; CURRENT APPLICATION NUMBER: US/10/640,833
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1168
; LENGTH: 5898
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-1168

Query Match
4.5%; Score 43; DB 6; Length 5898;
Best Local Similarity 48.6%; Pred. No. 0.093;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 408 ACCTGAATGACAGATGAAGCCTTAATGAAGCAAGCCTTGTGATGACACTCCAGATGTAAC 467
DB 1707 AGTTGAGATCAATAAATCAACAGAACATATGACTCTAAATAATGAAATTCAGAAAAAAC 1766
QY 468 GGCTCAATCATCCGCTTAATTAATGAAGTAAGGCCAAGAGCTCTGGAATAACCCAA 527
DB 1767 TGTAGAGAAAGTCCAGTAATCAATGAAGCAGTAGAGTAGTACTCAATCAATCAAG 1826
QY 528 GGCAGAGAGTGTGATTTGCTCAATTAAGCCAAAGATTAATCACTGATGAAAAACAA 587
DB 1827 AACAGAAAAACCAAGTTCACAGTAGAGTAAGAACAACTCTGGGAAATAGCTAA 1886
QY 588 AGAAATGCTGAGAAATTAATTAATTAATGAAGTAAGGCCAAGAGTCTGGAATAACCCAA 527
DB 1887 CGAAATGCTGAGAAATTAATTAATTAATGAAGTAAGGCCAAGAGTCTGGAATAAC 1946
QY 648 AAA 650
DB 1947 AAA 1949

RESULT 15
US-10-617-320-648
; Sequence 648, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 648:
SEQUENCE CHARACTERISTICS:
LENGTH: 5919 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...5919
SEQUENCE DESCRIPTION: SEQ ID NO: 648:
US-10-617-320-648

Query Match
4.5%; Score 43; DB 6; Length 5919;
Best Local Similarity 48.6%; Pred. No. 0.093;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 408 ACCTGAATGACAGATGAAGCCTTAATGAAGCAAGCCTTGTGATGACACTCCAGATGTAAC 467
DB 1728 AGTTGAGATCAATAAATCAACAGAACATATGACTCTAAATAATGAAATTCAGAAAAAAC 1787
QY 468 GGCTCAATCATCCGCTTAATTAATGAAGTAAGGCCAAGAGTCTGGAATAACCCAA 527
DB 1788 TGTAGAGAAAGTCCAGTAATCAATGAAGCAGTAGAGTAGTACTCAATCAATCAAG 1847
QY 528 GGCAGAGAGTGTGATTTGCTCAATTAAGCCAAAGATTAATCACTGATGAAAAACAA 587
DB 1848 AACAGAAAAACCAAGTTCACAGTAGAGTAAGAACAACTCTGGGAAATAGCTAA 1907
QY 588 AGAAATGCTGAGAAATTAATTAATTAATGAAGTAAGGCCAAGAGTCTGGAATAACCCAA 527
DB 1908 CGAAATGCTGAGAAATTAATTAATTAATGAAGTAAGGCCAAGAGTCTGGAATAAC 1967
QY 648 AAA 650
DB 1968 AAA 1970

Search completed: September 4, 2003, 21:56:06
Job time: 190 secs

```

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

|||||
b 241 KTRKSSNIDYKMLKTLVLTOKONDSTFVQSIGRELOANIKVDAFONIFROYIG 300
y 310 GGDSSSSSSSTNS 322
|||||
b 301 GGDSSSSSSSTNS 313

ESULT 4

AM5079

AAW5079 standard; Protein; 213 AA.

AAW5079;

02-OCT-1998 (first entry)

Streptococcus pneumoniae SP0021 protein.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
detection; pneumonia; otitis media; meningitis.

Streptococcus pneumoniae.

MO9818930-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19422.

31-OCT-1996; 96US-0029960.

(HOMA-) HUMAN GENOME SCI INC.

Choi GH, Hromocky J A, Johnson LS, Kunsch CA.

WPI: 1998-272224/24.

N-PSDB: AAV27340.

Nucleic acid encoding antigenic peptide(s) from Streptococcus
pneumoniae - or their epitope-containing fragments, useful in
protective or therapeutic vaccines, and for diagnosis

Claim 11; Page 55; 118pp; English.

The present sequence represents a protein from Streptococcus pneumoniae.
The nucleic acid sequence encoding the Streptococcus pneumoniae protein
can be useful in vaccines for inducing protective antibodies against
Streptococcus pneumoniae, for treatment or prevention of infection e.g.
pneumonia, otitis media or meningitis. Probes based on the nucleic acid
are used to detect Streptococcus infection (by usual hybridisation or
amplification methods), also for isolating Streptococcus genes or their
allelic variants. The protein can be used similarly to detect specific
antibodies in standard immunoassays, especially for diagnosing or
monitoring infections. Antibodies which bind the protein are used to
detect corresponding antigens, to purify the protein and for passive
immunisation (optionally coupled to a toxin). Vaccines are administered,
e.g. by injection, orally or through the skin, typically at 0.01-1000
µg (especially 10-300) µg/ml per dose.

Sequence 213 AA:

Query Match 57.7%; Score 916; DB 19; Length 213;
Best Local Similarity 93.6%; Pred. No. 3.3e-62;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKGSEAGADLISMGVITEHOFYEOVKSNSAQOVLNNTIKVFEKOGSELDKREVD 90
|||||
1 SKGSEAGADLISMGVITEHOFYEOVKSNSAQOVLNNTIKVFEKOGSELDKREVD 60
|||||
91 TIAEEKOYGENQVRLSQAQMTLETRKQIRTSKLVLAQKVAEALTDKAYKARDE 150
|||||
61 TIAEEKOYGENQVRLSQAQMTLETRKQIRTSKLVLAQKVAEALTDKAYKARDE 120

Fri Sep 5 09:24:50 2003

us-10-049-4

QY 151 YTPDYTAQIIRLNNEDKAKEVLEKAKAEGADFOALAKNDSTDEKRENGEITPDSASTE 210
|||||
Db 121 YTPDYTAQIIRLNNEDKAKEVLEKAKAEGADFOALAKNDSTDEKRENGEITPDSASTE 180
|||||
QY 211 VP-EQVKKAAFA-----LDYD 225
|||
Db 181 VPGASPKPLPAFRGCMFLDYD 203

RESULT 5

ABP54573

ABP54573 standard; Protein; 213 AA.

ABP54573;

04-SEP-2002 (first entry)

S. pneumoniae SP021 protein sequence SEQ ID NO:34.

Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
antibacterial; Streptococcal infection; detection.

Streptococcus pneumoniae.

US2002061545-A1.

03-MAY-2002.

22-JAN-2001; 2001US-0765272.

30-OCT-1997; 97US-0961083.

(CHOI/) CHOI G H.
(KUNS/) KUNSCH C A.
(BARA/) BARASH S C.
(DILL/) DILLON P J.
(DOUG/) DOUGHERTY B.
(FANN/) FANNON M R.
(ROSE/) ROSEN C A.

Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,
Rosen CA.WPI: 2002-479261/51.
N-PSDB: AB084808.New Streptococcus pneumoniae antigens, useful for detecting
Streptococcus and for preventing or attenuating disease caused by
Streptococcus infection

Claim 11; Page 24; 70pp; English.

AB084792 to AB084904 represents nucleic acids which encode the
Streptococcus pneumoniae antigens given in ABP54573 to ABP54659.
The S. pneumoniae antigens have antibacterial activity and can be
used in vaccines. The S. pneumoniae antigens can also be used to
prevent or attenuate a Streptococcal infection in an animal. The
polynucleotides encoding the S. pneumoniae antigens can be used to
detect Streptococcus nucleic acids. AB084905 to AB085130 represent
primers used in the cloning of S. pneumoniae ORFs (open reading frames)
which are used in an example from the present invention.

Sequence 213 AA:

Query Match 57.7%; Score 916; DB 23; Length 213;
Best Local Similarity 93.6%; Pred. No. 3.3e-62;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKGSEAGADLISMGVITEHOFYEOVKSNSAQOVLNNTIKVFEKOGSELDKREVD 90
|||||
1 SKGSEAGADLISMGVITEHOFYEOVKSNSAQOVLNNTIKVFEKOGSELDKREVD 60
|||||
91 TIAEEKOYGENQVRLSQAQMTLETRKQIRTSKLVLAQKVAEALTDKAYKARDE 150
|||||
61 TIAEEKOYGENQVRLSQAQMTLETRKQIRTSKLVLAQKVAEALTDKAYKARDE 120

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

57 09:24:50 2003

us-10-049-473a-2.ra1

Page 5

DB 70 TIOKVEK 77
DB 61 TIOKVEK 68
DB 130 AVKVAEALTEDEAKKAFDEYTPDYTAOIRLNNEDEKAEVEKAKAGADFAOLAKDN 189
DB 130 GLK--ANYKLTDKDIDTAMKEFEYPPVSTOILFSTKEEDAKKAEAN--EGENFSLVQAY 186
DB 190 STDEKTEKNGEITPDSASTVEPEQVKKAFALDVGSDVITATGQAYSQYIVLT 249
DB 187 GKMKLKTGDKMNDSTNPELPEYKKAAPKNGESDIPYDPTTYOOSYLVKXV 246
DB 250 KTEKSSNDYKKEKLTIVLTOKONSTVOSITIGKELQANINVKQAFONIF 304
DB 247 KQDGSNKKDKKSELEKATEKTLIDFEFMDKTIRKVKKNDVYTKIDPYKXNIF 301

LT 8
8-858-207A-508
Sequence 508, Application US/08858207A

Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328e1 Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Giumli, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090

TELEX:
INFORMATION FOR SEQ ID NO: 508:

SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 6348328e
US-08-858-207A-508

Query Match 20.3%; Score 322; DB 4; Length 74;
Best Local Similarity 97.1%; Pred. No. 1.2e-19;
Matches 66; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 10 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHGFYEVOKSNPSAQOVLMM 69
1 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHGFYEVOKSNPSAQOVLMM 60

Black et al.

DB 70 TIOKVEK 77
DB 61 TIOKVEK 68

RESULT 9
US-09-134-001C-3811
Sequence 3811, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3811
LENGTH: 330

TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3811

Query Match 13.2%; Score 210; DB 4; Length 330;
Best Local Similarity 27.5%; Pred. No. 1.4e-09;
Matches 95; Conservative 55; Mismatches 112; Indels 84; Gaps 20;

DB 23 SVATLAAC-SKGSAGAD--LISM-KGV-----ITEHGFYEVOKSNPSAQOVLMMIOY 74
DB 19 SALLGAGSNVAKESKDNLTLLISSAGDVKKADVKKKGRDI--AMTS-----FSIVLANKY 72
DB 75 FEKGYGSELDDKFEYDDTIAEKKOYG--ENQVRLSQAQMTLEKRAQIRTS--KLVEL 129
DB 73 LADTKKVDNDIDDKDKKEKQYGGKDPESMLKQGSMDIDYKEQKRLSANQKOLL 132
DB 130 AVKVAEALTEDEAKKAFDEYTPDYTAOIRLNNEDEKAEVEKAKAGADFAOLAKDN 178
DB 133 DKVNSDKIEKNSK-----TSHILIKVSKSDKSGSDKAKAEKIOKEV 182
DB 179 ---GADFAOLKDNSTDEKTEKNGEITPDSASTVEPEQVKKAFALDVGSDVI--TAT 234
DB 183 EKPNKFEJAKKESDSSSAKKDGLG--VKGQWVSFEKALFKLEGEYSKVKTIDY 241
DB 235 GQVAYSQYIVLTAKTE--KSSNI-----DDYKELKTIVLTOKO 274
DB 242 G-----THIKADKTEDFNSKSNIKOKLIEKVKPKLLTDVYKELLKRYVDK 294
DB 275 ND--STFOSITIGKELQANINVKQAFONIFQYIGGSSSSS 318
DB 295 RDIKKAIEDSLDPD-----KIKOOOQ-----OSOGGSGSLTNSGS 330

RESULT 10
US-09-252-991A-17422
Sequence 17422, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17422

DB 275 ND--STFOSITIGKELQANINVKQAFONIFQYIGGSSSSS 318
DB 295 RDIKKAIEDSLDPD-----KIKOOOQ-----OSOGGSGSLTNSGS 330

THIS PAGE BLANK (USPTO)

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)